DOCUMENT RESUME

ED 196 401 IR 008 845

AUTHOR Myrick, Howard A.; Keegan, Carol
TITLE Proceedings of the Technical Conference on Qualitative Television Ratings (1980). Final Report.
INSTITUTION Corporation for Public Broadcasting, Washington, D.C. Office of Communication Research.
PUB DATE 80
NOTE 101p.; Appendix D removed prior to filming.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Audiences; Bibliographies; *Evaluation Methods; Feedback; Models; *Programing (Broadcast); *Public Television; Quality Control; Rating Scales; *Research Methodology; *Television Research; Television Surveys; Television Viewing

ABSTRACT

The purpose of this conference was to assess the technical feasibility of obtaining qualitative ratings to measure the effectiveness of public television. The state-of-the-art in qualitative ratings is discussed in terms of a review of the Corporation for Public Broadcasting's past and current qualitative research efforts, international approaches to qualitative television ratings, new technologies and data collection alternatives, and methodological alternatives and considerations. Participants' recommendations for the design of a qualitative research system cover four principal areas: (1) its position in the total Office of Communication Research (OCR) program; (2) the nature of the data to be collected; (3) data collection and analysis; and (4) applications of research findings. The appendices include an overview of the international state-of-the-art in qualitative television ratings, a review study of several attitudinal models for the viewer evaluation of television programs, and a methodological assessment of existing evaluative ratings systems. (CHC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED196401

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PROCEEDINGS of the 1980

TECHNICAL CONFERENCE

on

QUALITATIVE TELEVISION RATINGS

Final Report

CORPORATION FOR PUBLIC BROADCASTING
OFFICE OF COMMUNICATION RESEARCH

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Corporation for Public Broadcasting

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

I200 P945

2

AUG 7 1980

PROCEEDINGS

CPB Technical Conference on Qualitative Television Ratings

(March 9-12, 1980)

Page

Preface

I. Introduction 1
   A. The Concept of Qualitative Television Ratings 1
   B. Purpose and Scope of the Conference 2

II. State-of-the-Art in Qualitative Television Ratings 6
   A. Review of CPB's Past and Current Qualitative Research Efforts 6
   B. International Approaches to Qualitative Television Ratings 15
   C. New Technologies and Data Collection Alternatives 20
   D. Methodological Alternatives and Considerations 22

III. Recommendations for the Design of a Qualitative Research System 28
   A. The Larger Research Context 28
   B. Nature of the Data to be Collected 29
   C. Data Collection and Analysis 30
   D. Applications of Research Findings 31

Footnotes and References 32

Appendices
   A. International State-of-the-Art in Qualitative Television Ratings
   B. Psychology of Viewer Program Evaluation
   C. Methodological Assessment of Existing Evaluative Ratings Systems
   D. List of Conference Participants

PREFACE

The Office of Communication Research is pleased to present the final report on the proceedings of its Technical Conference on Qualitative Television Ratings.

The search for a qualitative television ratings system is part of a larger search for programming evaluation criteria specially suited to public television. In spite of unique programming objectives in public television, we have failed to develop evaluation criteria and research procedures tailored to PTV programming philosophies. By default, we have generally fallen back on the audience research heritage developed by the commercial broadcasting sector.

The search for a qualitative television ratings system is also a recognition that at least some of these new program evaluation measures are rightfully obtained from a highly articulate viewing public which is able and ready to offer critical responses to television programming.

Finally, a qualitative television research program is, by its highly exploratory and rather irreverent nature, one which can question and revise conventional audience research practices. Given appropriate interpretation and application, it is our belief that qualitative research data can help us break free of the limitations of television as it is, to explore television as it might be.

Discussions at the Technical Conference confirmed our experience that the search for a qualitative television ratings system is no easy methodological undertaking. However, as this report evidences, a range of promising research approaches is immediately open to us, and the experiences of our predecessors in this area confirm the applicability of resulting data to varieties of television programming decisions.

We deeply appreciate the valuable contributions made by participants throughout the conference deliberations. Special thanks are extended to Herbert Dordick, James Anderson and Joseph Philport for their authorship of the three commissioned papers appended to this report. An additional acknowledgement is made to Leon Rosenbluth, former Director of the Office of Communication Research, who originated the Corporation's qualitative research program during his term here.

Howard A. Myrick
Director, Office of
Communication Research

Carol Keegan
Associate Director,
Office of Communication Research
Conference Manager

PROCEEDINGS

Corporation For Public Broadcasting (CPB)
Technical Conference on Qualitative Television Ratings
(March 9-12, 1980)

I. INTRODUCTION

A. The Concept of Qualitative Television Ratings

The search by CPB for a qualitative television ratings system grew from the belief that commercial television research practices yield inadequate measures of programming effectiveness for public television. Traditional measurements of audience size and composition were established to meet the information needs of the commercial television industry, where programs are ultimately aired, renewed or cancelled on the basis of their ability to draw audiences attractive to advertisers. Programming objectives in public television (PTV), however, do not constitute a simple search for mass audiences with marketable demographic characteristics.

Although the public broadcaster is also concerned about the size and composition of audiences, these concerns are mediated by a complex of programming objectives. For example, some programs are never intended to attract large audiences, but rather to serve smaller constituencies whose special needs and interests are not being addressed by other programming sources. Additionally, even where some PTV programs are marked for a general audience, they are often implicitly or explicitly intended to affect the viewer in some desirable educational, aesthetic, emotional or social sense. For these programs, drawing an audience is certainly a *sine qua non*, but merely the first in a set of programming objectives.

Although PTV has embraced special programming goals, it lacks parallel research activities with which to guide and evaluate institutional progress toward the achievement of those goals. The search for a comprehensive research program which addresses PTV's multiple programming objectives has led CPB into the area of qualitative television audience research.

1. Defining Qualitative Ratings. In general, proponents of qualitative television research advocate the establishment of an expanded viewer feedback channel.

Currently, the viewer has only a limited, nonverbal vocabulary with which to comment on the value of various television programs. Viewers can make binary comments on each program offering, either by tuning in or by not tuning in. This limited vocabulary leaves the viewer highly inarticulate as a program critic, since many factors which influence viewing behavior have little or nothing to do with a critical evaluation of program offerings [5].

A qualitative ratings system, then, would endow viewers with a new program evaluation vocabulary, and a feedback channel through which comments can be delivered to those persons and organizations which originate programming. These new program ratings are not intended to replace or challenge traditional ratings. Rather, they would provide entirely different kinds of information which reveal the extent to which program offerings provide worthwhile viewing experiences to their audiences.

Finally, to the extent that this new research can diagnostically pinpoint the sources of programming success and failure (in both viewer-defined and producer-defined terms), a qualitative television ratings system should ultimately improve the development of future programming.

Unfortunately, the principal proponents of qualitative television ratings (qualitative research providers, media reform groups, foundations) have not extended the definition of qualitative television research much further. Customarily, one is told that qualitative program ratings will supplement traditional measures of audience size and composition and tell us "something more" about the nature of the viewer's response to programming. Usually undefined in such statements, however, are the specific qualitative research variables, the specific program evaluation criteria which the new ratings system will measure.

Defining such qualitative research variables is particularly challenging for PTV where program originators often strive to achieve higher-order programming objectives. Programs are designed to convey information, to stimulate additional information-seeking, to develop viewers' tastes and interests, and enhance personal and social coping skills. One can envision a continuum of implicitly and explicitly intended impacts for public television programming.

Conceptually, qualitative television ratings could be mapped onto this continuum of viewing impacts. The domain of qualitative ratings information for PTV would pick up where traditional ratings data stop. Theoretically, qualitative ratings variables could cover the complete range of viewing impact variables. (See Figure 1, p. 3.)

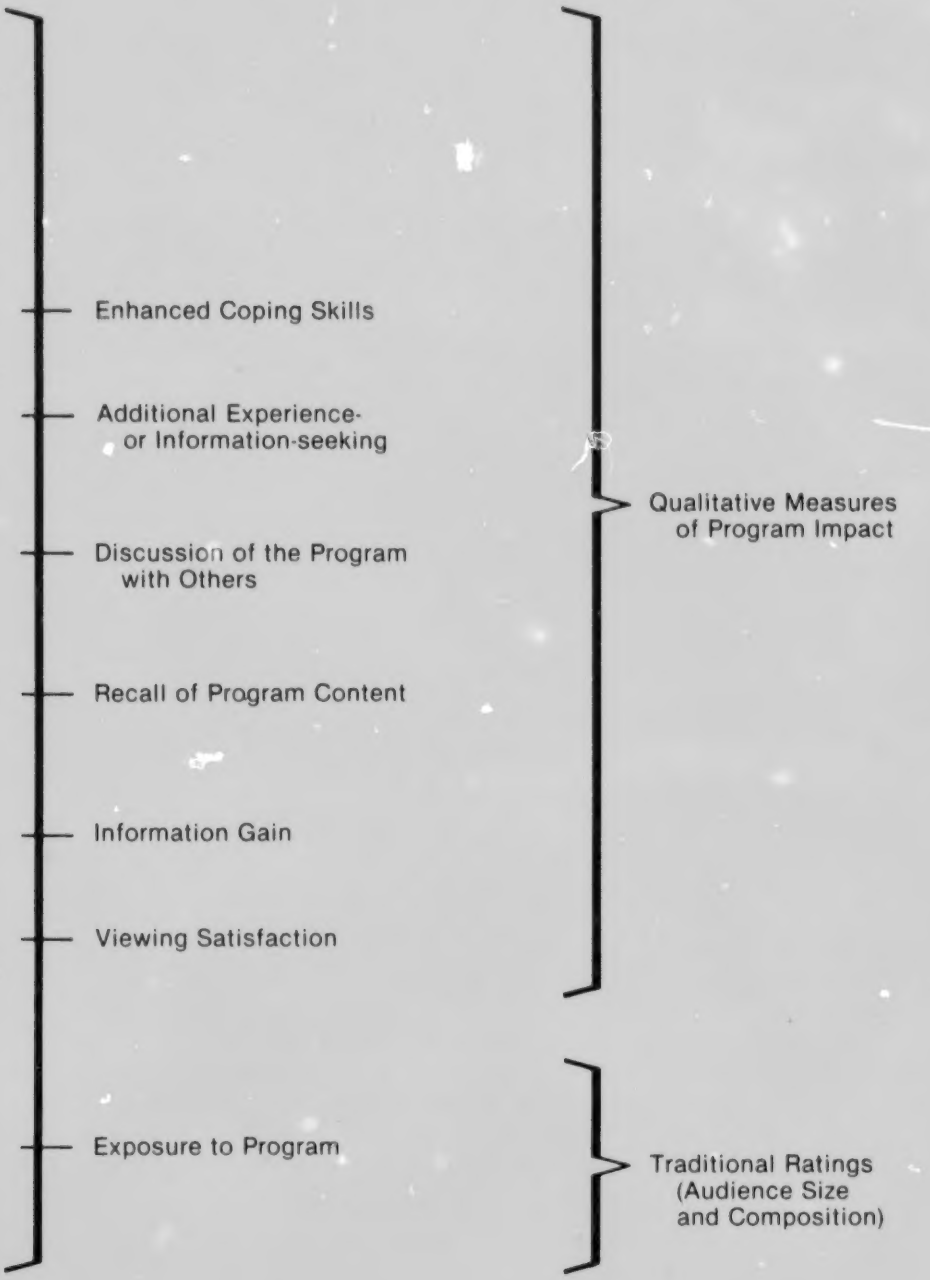
Most existing qualitative ratings systems, however, focus on the lower end of the qualitative research continuum. Primarily, they measure viewing satisfaction, the extent to which viewers find a program appealing or enjoyable [1,3]. If qualitative television research is to address questions of program effectiveness in terms of additional evaluation criteria, it appears that PTV, not commercial television, will have to provide the initiative.

B. Purpose and Scope of the Conference

To assess the technical feasiblility of obtaining qualitative ratings which would speak to such concerns, CPB's Office of Communication Research (OCR) convened a special conference from March 9-12, 1980. The conference assembled users, providers and proponents of qualitative television research, as well as commercial and

Figure 1.

Parallel Viewing Impact and Research Continua



academic communication researchers.[1]

Conference participants were fully briefed on OCR's past research with existing qualitative ratings systems, and on its future research plans in this area. Specifically, participants were asked to assess the technical feasibility of near-term development of a qualitative ratings system attuned to PTV's information needs. Additionally, they were asked to recommend criteria for assessing the cost-effectiveness of various qualitative television research practices.

Given its previous experience with several qualitative research projects, OCR specified some desirable features of a qualitative ratings system. An effective qualitative ratings system should:

- Not attempt to replace or challenge traditional program ratings; rather it would explore a range of entirely different viewing impacts. Ideally, viewing impact variables covered in the new ratings would reflect special PTV programming goals and objectives.
- Obtain comparable viewer responses to all nationally broadcast television programs, with the same qualitative research measurements applied to all programs.[2]
- Be developed with great concern for scientific rigor in data collection and analysis procedures. Data should be generalizable to the total population of television viewers, as well as to significant population subgroups.
- Not limit itself to the qualitative evaluation of PTV programs. It should collect comparable data for public *and* commercial television programs. Public television viewing occurs within the viewer's larger familiarity with the medium of television and all its offerings. If we are to understand responses to public television programs, we must analyze both commercial and public viewing experiences.
- Be ongoing. At least several times a year, we should obtain "other numbers"—qualitative indices to set side-by-side with traditional audience ratings for better informed evaluations of programming effectiveness.
- Be analyzed and reported in a manner which maximizes their relevance in a variety of decision-making arenas in PTV programming. Across the different levels of programming responsibility in PTV, qualitative ratings should be useful in guiding decisions about program planning, funding, scheduling, promotion and evaluation.

With regard to this last feature, participants discussed the potential applications of qualitative television ratings for three sets of PTV programming decision-makers: the CPB Board of Directors, CPB's Program Fund and local (PTV) stations.

1. The CPB Board of Directors. This governing body is ultimately responsible for guiding the Corporation's expenditure of funds in support of PTV programming. Conference participants were asked to consider whether qualitative research could be designed to assist CPB in setting its program priorities, and in evaluating corporate and institutional progress in implementing them:

- Could qualitative ratings help the Board operationally define programming

objectives like providing "alternative" programming and fostering "diversity?" Could the new research help the Board specify the implications of these goals for programming decision-makers in the Fund?

- Could it help define "alternative" programming as material not available to viewers from other sources, as viewing needs and interests which research indicates are not being addressed by other programming originators?
- Can qualitative research help generate new and diverse programming directions which will speak to a variety of unsatisfied viewer programming preferences?
- Can qualitative ratings provide evaluation data parallel to such programming directions, so that the Board can be appraised of the consequences of corporate programming investments made through the Fund?
- Most importantly, can these evaluation data surpass traditional audience ratings, to explore success in attaining special PTV programming objectives (the extent to which an audience was moved, educated, stimulated to gather additional information, introduced to new aspects of a social or political issue) [15]?

2. *The CPB Program Fund.* The Fund is the locus of all program funding decisions within the Corporation. It is responsible for the pursuit of CPB Board-approved program policies, priorities and procedures, although the Fund operates independently in matters of individual program funding decisions. The Program Fund is in a unique position to apply qualitative ratings data to program development and evaluation activities.

With regard to program development, OCR is exploring whether qualitative research data can be used to identify programming gaps currently unaddressed by public and commercial television. Specifically, can qualitative data be used to:

- Compare institutional and program-specific images for PTV and CTV and suggest "alternative" directions for PTV—directions not being undertaken by other programming sources?
- Identify target groups not generally attracted to, or satisfied with PTV programming, identify their viewing preferences and suggest program concepts and treatments sensitive to their special concerns?
- Fine-tune promotional work for targeted programming, based on known target group program perceptions and preferences?

With regard to program evaluation, OCR is working with the Program Fund to apply qualitative ratings data to a better understanding of traditional program ratings data. These analyses are conducted to assist with Fund decisions to continue or withdraw CPB support for programs. Several potential applications of the data are being studied:

- Can public awareness of a program be analyzed to determine whether programs are suffering from inadequate promotion?
- Can preference data for competing commercial programs be used to assess the impact of national feed and local scheduling practices on audience size?

- Can targeted programs be judged not merely on total audience size and satisfaction, but in terms of the reactions of those viewers for whom the program is intended?

3. Local PTV Stations. In addition to exploring qualitative research data for potential applications to internal corporate decision-making and for public and Congressional accountability, CPB has been exploring applications of qualitative ratings data for local PTV station program managers and researchers. Application areas being explored include the following:

- Would these data be useful to the system in day-to-day program carriage, scheduling and promotion decisions?
- Could they be used to enhance audience flow, to avoid problematic scheduling competition or to attract underserved viewer subgroups?

The potential relevance of qualitative research to these three varieties of programming decisions will be the focus of subsequent CPB conferences and workshops. Methodologists and research practitioners at the March technical conference were directed to consider these application areas as they approached various research system design considerations.

II. STATE-OF-THE-ART IN QUALITATIVE TELEVISION RATINGS

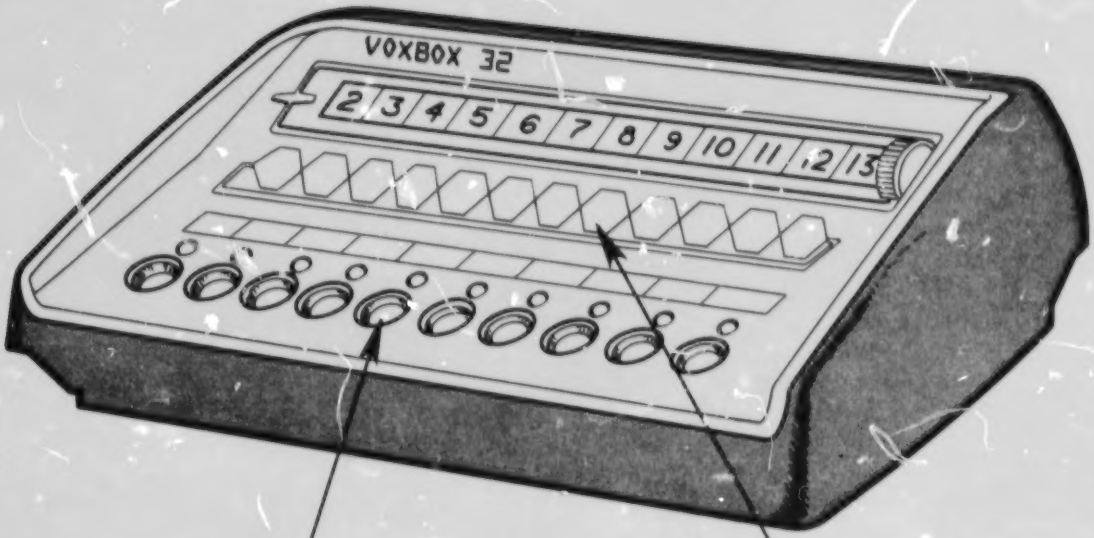
Conference proceedings began with discussions of existing qualitative ratings systems and a review of existing domestic systems. (See Appendix C, p. 2.) Early conference sessions were devoted to in-depth discussion of OCR experiments with three of these systems: the Percy System, PTvQ and TQR.

A. Review of CPB's Past and Current Qualitative Research Efforts

1. The Percy System. An electronic response device called the VOXBOX is the basis of the Seattle-based Percy System. The Percy panel is composed of 200 television households in which a VOXBOX has been wired to the main television set. Panelists agree to record their spontaneous evaluations of whatever television programs they choose to watch, using eight specially labeled qualitative response buttons on the VOXBOX. (See Figure 2, p. 7.) Channel selections and qualitative responses are fed continuously into the system's main computer via special phone lines. Data are collected 24 hours a day, seven days a week. Over the next few years, the Percy Company plans to set up similar panels in additional markets.

Applications of the Data. Throughout 1979 CPB subscribed to the Percy System on a trial basis and received both syndicated data and reports specially requested by OCR. The limited representativeness of the Percy panel, combined with a low incidence of PTV viewing (and accordingly, a low incidence of qualitative responses to PTV programs) made the syndicated data of limited use. However, special studies made possible through a "testbed" relationship with KCTS, Seattle have shown promise for formative research.

Figure 2
Percy System VOXBOX



Qualitative Response Buttons

Channel Selection Buttons

"Excellent"
"Informative"
"Credible"
"Funny"
"Boring"
"Unbelievable"
"Dumb"
"Zap"
"Person" - special button used to apply a qualitative response to a program personality, rather than to program content

"D" - special button used to identify specific household member using the VOXBOX

Pilot testing, for example, can be greatly elaborated through the Percy System. In the past, most CPB studies of television pilots have relied on small focus group discussions, in which a group leader shows a pilot to 20 or 30 individuals who gather at a special test site to discuss the program's relative appeal. Such procedures are suspect with regard to the small number of viewers whose reactions are solicited, artificial viewing conditions and the possibility of group contagion of opinion. The Percy System, however, provided OCR with an interesting methodological alternative for pilot testing. By recruiting Percy panelists to watch a pilot program specially aired by KCTS, opinions of many more viewers can be elicited. Furthermore, these individuals use a standardized vocabulary (i.e., the button labels) and make judgments in the private and "normal" conditions of home viewing.

In one specific pilot test, during which 140 Percy panel households cooperated with the recruitment, OCR explored the potential of the system for providing formative research data to the producer of a pilot program. Recruitment instructions were developed to allow viewers their usual "free-running" responses to the pilot, but we also elicited responses (via a small flashing signal in the lower left corner of the screen) at several points in the program where the producer indicated he was most curious about the audience's response [7].

In another formative application, KCTS' production staff used the Percy System to fine-tune the development of a news series. Producers and researchers collaborated closely in applying the Percy panel data over the eight-week period of KCTS's broadcast of the series. Throughout these weeks, production adjustments were based on VOXBOX and supplementary research data and were followed by both increased audience size and increased positivity of viewer response [12].

Other special studies are anticipated as OCR moves from the subscription of syndicated VOXBOX data to a variety of customized research projects. In cooperation with KCTS, OCR plans to explore the effects of selected fund-raising, program promotion and scheduling practices on both viewing behavior and qualitative response.

2. TvQ Service. A second CPB experiment with an existing qualitative ratings system involved the development of a public television version of TvQ, a national mail survey of television viewers.

For almost 20 years the TvQ service has provided commercial industry decision-makers with: first, measures of national awareness of virtually all current commercial programming, and second, viewers' overall program evaluations along a five-point scale.

In cooperation with Marketing Evaluations, Inc., providers of the TvQ service, TvQ research procedures were adapted to develop a new "PTvQ" service. Through prescreening surveys, a national pool of PTV households was developed. In each of two survey waves, 1200 PTV viewers were drawn from this pool and asked to indicate whether they were "familiar with" a sample of approximately 100 PTV programs. If familiar, they were then asked to rate each program's appeal along a five-point scale, ranging from "one of my favorites" to "poor." As in the TvQ service,

two basic scores were derived: a familiarity score (percentage of respondents indicating familiarity with a program) and a PTvQ score (percentage of respondents familiar with a program who claim it is one of their favorites).

For the same PTV viewer samples, comparable scores were available (through coincident TvQ survey waves) for virtually all commercial television programs. Additionally, for the two special PTvQ waves, Black and Hispanic respondents were oversampled, in order to permit analysis of the program preferences of these audiences.

Applications of the Data. The principal purpose of these PTvQ studies was to assess the feasibility of collecting an overall qualitative program rating for PTV programs, similar to those collected by British ("appreciation index"), Canadian ("enjoyment index"), and French ("interest index") broadcasters [3]. The two survey waves were successful explorations, in terms of identifying effective PTV viewer sampling procedures, obtaining a range of apparently discriminating viewer program evaluations, and uncovering provocative demographic variations in audience response.

Some of the most interesting research questions studied with these data were whether PTV viewers appeared satisfied by available offerings, and whether they appreciated PTV programs more than commercial offerings. Table 1 compares average Q scores for public and commercial programs studied in the two PTvQ surveys conducted to date.

Table 1

Q Scores for Public vs. Commercial Television Programs
(PTV Viewer Samples)

	WAVE I (Oct.-Nov. 1978)			WAVE II (July-Aug. 1979)	
	PTvQ	TvQ		PTvQ	TvQ
Average Score =	25	26	Average Score =	26	22

Average qualitative evaluations indicated that PTV viewers were no more appreciative of public television programs than they were of commercial television programs, although certain public television programs far exceeded average qualitative evaluations.

Table 2

Top Ten PTvQ Programs

WAVE I (Mean PTvQ Score = 25)		Q
1.	*Evening at Pops*	45
2.	*Mark Russell*	43
3.3	*Evening at Symphony*	40
3.3	*Nova*	40
3.3	*Cousteau:* "Oasis in Space"	40
6.	*Adams Chronicles*	38
7.5	*Wall Street Week*	37
7.5	*Crockett's Victory Garden*	37
9.	*Masterpiece Theatre:* "Duchess of Duke Street"	36
10.	*Pompeii Frozen in Fire*	35

WAVE II (Mean PTvQ Score = 26)		Q
1.	*Live from Lincoln Center:* "Zubin Mehta and Itzhak Perlman"	54
2.	*National Geographic Special:* "Gold"	51
3.	"Cousteau Odyssey"	50
4.5	"Arthur Fiedler Call Me Maestro"	49
4.5	*National Geographic Special:* "Tigris Expedition"	49
6.5	"Horowitz Live Encore"	47
6.5	"Treasures of Tutankhamun"	47
8.	*Masterpiece Theatre:* "Lillie"	46
9.	"Baryshnikov at the White House"	43
10.5	*Masterpiece Theatre:* "Duchess of Duke Street"	42
10.5	*Live from Lincoln Center:* "Sleeping Beauty"	42

Program "familiarity" scores (FAM scores) produced provocative results, suggesting an important area for potential applications of PTvQ data. In WAVE I, even among these samples of PTV viewers, familiarity with public television programs was much less common (average FAM=18) than familiarity with commercial television programs (average FAM=64). The performance series was the most familiar program type (average FAM=34); performance specials and documentaries were the least well known (average FAM=8). Among the best known programs were: *Sesame Street* (FAM=64), *The Electric Company* (FAM=55), *Dick Cavett* (FAM=50) and

Mister Rogers' Neighborhood (FAM=50) [13]. Valuable applications of the PTvQ data, then, might include tracking awareness of individual public television programs, and providing guidance for institutional and individual program promotion activities.

Results of these PTvQ surveys underscore the importance of using special evaluation practices for targeted programs. Programs intended for a minority or special interest audience may not fare well with the total viewing audience, yet will apparently satisfy special needs and interests among their smaller viewership. With PTvQ data, such differential audience reactions can be measured for more informed decisions about the effectiveness of targeted programming. With Q-like data, program funders and schedulers could identify and support programs which fail to attract large audiences, but which prove satisfying to the smaller audience they do attract.

For example, PTvQ data provide interesting findings with which to evaluate *Over Easy*, a PTV program targeted to viewers 55 years of age and older. In both waves, PTvQ data showed high target group satisfaction with *Over Easy*, while the total sample evaluated it below average. They also revealed an unfortunately low level of target group awareness of the program.

Table 3

Comparison of Target Group and
General Audience Response to *Over Easy*

	WAVE I		WAVE II	
	Target Group (55+)	*Total Sample*	*Target Group*	*Total Sample*
PTvQ:	39	20	39	29
FAM:	16	16	21	29

The *Over Easy* case also pointed out some potential applications of the TvQ data as input to program scheduling decisions. The program's target group (persons 55 years of age and older) evaluated commercial news programs very favorably. (See Table 4, p. 12.)

Obviously *Over Easy* (or any other program targeted to this audience) would suffer from strong competition if aired in time slots occupied by commercial news programming.

While the Percy System provides a small sample of viewers making numerous and

varied judgments about their responses to television programs, the PTvQ service delivers a much larger and demographically varied sample making very simple program evaluations. The Corporation's third major qualitative research project sought to combine the more desirable features of both these systems, through the newly developed TQR Service. Through TQR, multi-dimensional viewer evaluations were obtained from a national sample of respondents.

Table 4

Top Ten TvQ Programs
(Respondents 55 Years of Age and Older)

WAVE I			WAVE II		
1.	*60 Minutes*	65	1.	*60 Minutes*	66
2.	*CBS Evening News with Walter Cronkite*	53	2.	*CBS Evening News with Walter Cronkite*	50
3.	*NBC Nightly News with Chancellor & Brinkley*	48	3.	*CBS Sunday News with Ed Bradley*	45
4.	*M*A*S*H*	46	4.	*M*A*S*H*	43
5.	*Diff'rent Strokes*	43	5.	*CBS Newsbreak*	42
6.	*Quincy*	42	6.3	*ABC News Weekend Report (Sunday)*	41
7.	*The Waltons*	41	6.3	*ABC News Weekend Report (Saturday)*	41
8.	*NBC News Update*	40	6.3	*CBS Evening News with Morton Dean (Sunday)*	41
9.5	*Alice*	39	9.	*NBC Nightly News with Chancellor & Brinkley*	40
9.5	*Lifeline*	39	10.3	*Diff'rent Strokes*	39
			10.3	*ABC News Brief*	39
			10.3	*ABC News World News Tonight*	39

3. Television Qualitative Ratings (TQR)

a. Description of the System. The TQR Service provides a profile of television programs along 14 factors derived from an extensive analysis of viewer program perceptions. In a pilot study, conducted by Marketing Evaluations, Inc., 258 descriptors of television programs (as identified in preliminary focus group discussions and analyses of the television criticism literature) were used to profile 96 commercial and

public television programs. Judgments as to whether each descriptor did or did not apply to all 96 programs were collected from a total national sample of 3,000 respondents. Data reduction yielded the identification of 14 factors in viewer program evaluations. As a syndicated service, TQR will provide ongoing measurement of current programming along these 14 factors, and along the complete set of descriptors from which the factors emerged:

What the Program Does for the Viewer

Factor 1—Provides KNOWLEDGE AND ENRICHMENT
Factor 2—Provides DIVERSION AND ESCAPE

How the Program Affects the Viewer

Factor 3—Gives FUN AND AMUSEMENT
Factor 4—Builds TENSION AND EXCITEMENT

How the Viewer Feels about the Program

Factor 5—POSITIVE EVALUATION - wants to watch
Factor 6—NEGATIVE EVALUATION - not interested

The Program's Type of Appeal

Factor 7—Acceptable for KID/FAMILY VIEWING
Factor 8—Primarily for ADULT ENTERTAINMENT

The Contents of the Program

Factor 9—Subject/People are FAMILIAR AND REALISTIC
Factor 10—Subject/People are UNFAMILIAR OR UNUSUAL

The Viewer Relationship to the Program

Factor 11—Viewer becomes EMOTIONALLY INVOLVED
Factor 12—Viewer is INTERESTED OBSERVER

Sensory Emphasis of the Program

Factor 13—Visual stress on BEAUTY AND GLAMOUR
Factor 14—Aural stress on WIT AND CLEVER LINES

As a supplement to TvQ, TQR is intended to provide diagnostic information about the sources of viewer (dis)satisfaction with a given television program. The CPB Office of Communication Research co-sponsored the factor analysis pilot study, and subscribed to the first TQR survey conducted in November and December 1979.

b. *Interpretation of the Data.* One of the first questions asked of the data was whether PTV was seen as providing a certain type of viewing material, whether the image of PTV programming appeared to be focused on any one or a few program evaluation factors. Data from this first TQR survey showed PTV programs loading most heavily on the Knowledge-Enrichment Factor. (See Table 5, p. 14.)

To date, secondary analyses of the TQR data have concentrated on the identification of "alternative" programming areas; that is, gaps in available programming which are not being addressed by commercial or public sources. One cognitive mapping of viewers' perceptual space for television programming, for example, identified "holes" in available offerings where PTV programs could be developed to

satisfy unaddressed viewing interests, and in so doing, to diversify the programming available to the viewing population [8].

Table 5

Twenty Programs with Highest Factor Loadings
on
"Knowledge and Enrichment"

	Provides Knowledge and Enrichment
60 Minutes	62
20/20	52
**Nova*	51
CBS Evening News with Walter Cronkite	50
Good Morning, America	47
**MacNeil/Lehrer Report*	46
**Wall Street Week*	44
NBC Nightly News with John Chancellor	44
Phil Donahue	43
Issues and Answers	42
**Bill Moyers' Journal*	41
Today	40
Tomorrow (with Tom Snyder)	39
Prime Time Sunday	39
Meet the Press	37
**Washington Week in Review*	36
The Bible. . . in the Beginning	30
**Firing Line*	29
ABC Afterschool Specials	29
Wild Kingdom	29

*PTV Program

Future analyses will explore the utility of TQR data as formative input to the development of targeted programming. In combination with traditional program ratings, can TQR data be used to identify content areas, formats, treatments, and other production elements associated with a target group's favorable and unfavora-

ble reactions to programming? What would such analyses suggest for optimal promotion and scheduling of programs targeted to a given audience?

B. International Approaches to Qualitative Television Ratings

The United States has a great deal to learn from abroad. A number of other countries have had considerably earlier and more influential roles in the development of innovative audience research methodologies. Appendix A, a paper specially commissioned for the conference, is an overview of qualitative television ratings systems in several other nations.

Sweden, the Netherlands, France, Britain, Japan and Canada exemplify the most advanced approaches. It should be noted, that unlike the United States, each of these nations has a public (i.e., government owned and operated) broadcasting system. These countries exhibit a tendency for qualitative ratings to thrive when broadcasting structures are under the greatest public and governmental pressure to justify their continued funding and operation.

The paper also reveals certain commonalities in approach which speak to larger issues about the design and operation of a qualitative ratings system.

On a simply methodological level, there are convergences across the six nations in their reliance on traditional survey research data collection channels (telephone interview, personal interview, diaries) and in the measurement of an overall qualitative rating score or index.

Approaches diverge, however, when one examines the total qualitative research program in which summary qualitative ratings are embedded.

In each nation studied, the overall ratings scale, whether based on "enjoyment" (Canada), "interest" (France) or "appreciation" (British), is supplemented by a battery of related and much more extensive research on viewer needs and preferences. These supplementary qualitative research efforts vary considerably. They range from formative studies designed to generate new programming ideas and concepts to special summative evaluations of individual programs; from national ascertainment studies of public needs and interests, to diagnostic research on a select number of broadcast programs; and from psychographic studies of popular values, to 24-hour a day time-budgeting studies. In some cases, overall institutional images for a programming source are qualitatively rated as thoroughly as individual programs are analyzed.

Across the countries studied, qualitative research appears to be a total package of research efforts uniquely addressed to the special information needs of individual broadcast entities. Approaches to research design and analysis area as individualistic as the socio-political environment in which each broadcasting institution operates.

A second important observation is the close interrelationship between traditional program ratings and qualitative research data. In the countries surveyed, traditional measures of audience size and composition are frequently collected in the same diary or interview used to obtain qualitative ratings. Furthermore, the two program

ratings are often co-reported, with their interrelationships analyzed (e.g., correlations between audience size and overall qualitative measures of program appeal). Table 6 presents a typical reporting format for Canadian audience research data, with traditional and qualitative program ratings presented side by side. Additionally, some research systems explore a "borderline" ratings area, the secondary analysis of traditional program ratings for audience measures which appear to indicate viewing satisfaction (e.g., repeat viewing, channel loyalty).

Representatives of the Audience Research Center, French National Broadcasting System, attended the conference and reviewed the history of qualitative ratings in France, analyzing some of the principal challenges facing the introduction and continued operation of such research efforts.

Since 1967, French audience research has counterbalanced measures of audience size with a qualitative viewer "interest" score, supplied both by an ongoing diary panel and special national surveys. The impetus for such research came directly from broadcasting legislation, which in 1974 charged the broadcasting system with providing programming to satisfy multiple objectives. Both radio and television channels were directed to foster variety, originality and innovation in their programming; to strive for high levels of artistic and cultural merit; and "to meet the expectations of different segments of the public...."

The current qualitative television research program was born in response to these mandates.[3] Most interesting is its primary focus on the television audience as the source of measurements for these various accountability criteria. Viewer perceptions of program quality (in terms of variety, innovation, artistic merit, etc.) are studied to assess programming effectiveness. These criteria are not reserved for "expert" judgment by panels of critics, producers or social commentators. Ultimate accountability rests in the judgments of the viewing population.

The Audience Research Center encountered an essential decision about three varying approaches to program "quality"; approaches which would dictate three highly different evaluation research methodologies.

One approach argues that a good program is one which conforms to classic aesthetic and cultural values. Since these values are supposedly independent of the vagaries of public opinion and mass culture, this conception would dictate a research program that focused on the opinions of a small number of "experts" and critics.

A second approach sees television programs as purposive vehicles for achieving some intended effect on viewers (e.g., conveying specific information to a target audience). The effectiveness of programs, then, would be measured by comparing viewing impacts intended by program originators with actual viewing consequences. A qualitative research program based on this approach would concentrate on getting program originators to articulate measurable programming goals for individual programs, and conducting special summative evaluations of the effectiveness of those programs in achieving these goals.

Table 6

Enjoyment and Audience Composition
(Average: Sept. to Dec. 1979)

		All (2+)	All (12+)	Young (2-11)	Young (12-17)	Adults (18-36)	Adults (35-49)	Adults (50+)	Men 18+	Women 18+	
Population[1]	%	100	84	16	13	32	16	23	35	36	
Archie Bunker	%	100	91	10	7	20	14	50	39	45	Monday
1,491,000	E.I.	N.A.	71	N.A.	66	66	69	75	72	71	8:00-8:30
F.D. Challenge	%	100	96	4	2	14	12	68	41	53	Friday
1,438,000	E.I.	N.A.	81	N.A.	IFR	64	76	86	78	85	8:30-9:00
The RAES	%	100	88	12	18	34	13	24	29	41	Tuesday
1,119,000	E.I.	N.A.	59	N.A.	688	60	56	51	54	58	8:30-9:00
Mork and Mindy	%	100	80	20	15	33	14	18	31	34	Sunday
2,469,000	E.I.	N.A.	74	N.A.	83	75	66	62	70	71	7:30-8:00

Source: Canadian Broadcasting Corporation, English Services Division, Research Services, Toronto.

Finally, the third approach to program "quality" would posit all judgments of programming effectiveness in the minds of the viewing public. A "good" program would be one which satisfied its audience. Accordingly, a qualitative television research program would principally measure program appeal.

Judging that the 1974 broadcasting legislation required measurements of programming effectiveness related to all three approaches, the Audience Research Center proceeded to design a qualitative research program which addressed the quality criteria implied in all three approaches, gathering all related evaluation information from viewers' judgments. They designed a research system based on nine dimensions of quality (see Figure 3, p. 18), four of which address evaluation criteria derived from the first approach (cultural value, originality of form, originality of subject matter, production quality); three of which tap possible pro-social effects of viewing (informativeness, increasing understanding of current events, coverage of different population subgroups); and two of which measure viewers' liking or disliking of programs (entertainment value, interesting subject matter).

Figure 3

The Nine Dimensions of Quality

Audience Research Center
French National Broadcasting System

Dimensions	*Criteria*
— Interest in the subject matter	— "A program whose subject interests you"
— Interpretation of events	— "A program which enables you to understand the world in which we live"
— Today's culture	— "A program which brings your knowledge up-to-date"
— Traditional culture	— "A program with cultural value"
— Entertainment value	— "A program which is enjoyable"
— Coverage	— "A program for everyone"
— Originality of the subject	— "A program whose subject is out of the ordinary"
— Originality of form	— "A program done in an original manner"
— Production quality	— "A well done program"

Early system development efforts were similar to the TQR research. Focus group studies were conducted to study viewer vocabulary vis-a-vis television programs. Eighteen of the 76 program descriptors generated in this manner were selected to address accountability criteria suggested by the legislation, and to base program evaluations in viewer-defined criteria for "good" television. These 18 criteria were then taken to a national sample of 4,230 viewers who were asked whether or not each applied to a sample of programs.

This experiment underscored the financial burdens inherent in performing such research on a continuing basis. It also revealed the "inappropriateness" of some of the criteria when viewers tried to apply them to certain programs. In 1976, the research design was revised in three respects.

- Nine criteria (presented in Figure 3) were selected from the original 18, to represent a balanced measurement of viewer programming preferences and the concerns of public authorities for the cultural and socially purposive aspects of programming.

- A provision was made for weighting the nine criteria in terms of public salience. Viewers were asked to report how important each criterion was to them in terms of television viewing.

- Research cost problems were addressed by simplifying the qualitative measurements of the nine criteria enough to allow their inclusion in traditional viewing diaries maintained by a national panel. More detailed qualitative measurements along the same nine dimensions would be obtained from two or three special studies a year.

Accordingly, the qualitative television research program in France today rests on two interrelated data collection procedures. A national panel of viewers submits both traditional program ratings data (viewing by quarter hour, throughout the broadcast day, for seven days a week), and an overall qualitative program rating for viewed programs (on a six-point scale measuring degree of "interest" in the program). Additional qualitative ratings measure salience of the nine evaluation criteria and ask panelists whether or not each applies to two recently viewed programs selected by the respondent as "informative," two selected as "entertaining," and two as "culturally enriching." These diary measures bring in qualitative evaluations from 1,000 panelists per week, approximately 50,000 per year.

In two or three supplementary field surveys annually, 42,000 interviews also measure personal salience of the nine criteria and viewer evaluations of five recently viewed programs along nine related bipolar scales.

Finally, qualitative program ratings are derived from the two data sources, weighting each criterion for public salience and each program for length. Monthly, qualitative ratings are reported to each television channel, including one overall rating and channel ratings along each of the nine criteria. These ratings are reported both for channel programming *in toto*, and for each program type. Yearly, performance data

for each television channel are provided to public authorities to inform their allocation of tax revenues.

The French researchers concluded that three closely interrelated problems confront the development of a new qualitative television ratings system: first, defining the qualitative evaluation criteria to be assessed; second, measuring performance in a manner that meets the information requirements of different users (e.g., legislators vs. broadcasters); and third, identifying cost-effective data collection procedures [4].

C. New Technologies and Data Collection Alternatives

The current state-of-the-art in qualitative television ratings was extended through conference presentations on two nontraditional data collection channels which might prove suitable for qualitative research efforts. The two new data collection channels discussed in this context were: PEAC, a portable program analyzer; and PAL, a developing system which relies on touch-tone telephones for data entry from the home.

1. PAL, Consumers' Computer. Product and Area Locator (PAL), developed by the Consumers' Computer Corporation of America, is an interactive system which uses touch-tone telephones to give individuals in major markets around the country at-home access to a variety of computerized consumer services. Currently, the system is being field-tested in San Diego, with expansion to major markets across the country anticipated by the end of 1980.

The system provides a new channel for public opinion research. Any individual with a touch-tone phone can participate in survey research efforts, registering responses with the 12 keys on the touch-tone pad. Questions can be presented by tape-recordings used during the actual interview (with the computer pacing and branching the presentation of interview items contingent on respondent tone entries), or through prior instruction via newspaper ad, special mailing or operating instructions for a panel arrangement.

Such capabilities suggest that it may be possible to devise a nationwide qualitative ratings panel, using touch-tone telephones for direct data entry by viewers. Panelists might be asked to use their touch-tone capabilities to record overall program evaluations immediately after viewing. In individual markets, data could easily be aggregated overnight, with cross-market summaries available shortly thereafter. Panelists might be instructed to view programs of their own choosing, then to register program evaluations along several program rating scales. Time intervals between viewing and evaluation could be controlled by instructing panelists to evaluate a program immediately after viewing or within a minimum time period, and by programming the computer to reject data not acceptably time-coded.

The entire procedure, including entry of a panelist identification code and program codes should take less than one minute. Panelists could either be cued by PAL and verbally guided through the several questionnaire items, or they could operate from prior instructions. Additionally, special studies might be conducted, recruiting panelists to view a specific program and to provide more extensive responses.

Unlike the Percy System, panel membership is easily rotated, without the costly and intrusive requirement of installing special hardware. However, PAL does not have the Percy System capacity for electronic verification of panelist viewing of evaluated programs.

In June 1980, the Corporation will field test this potential data collection channel for qualitative ratings, testing the system's capacity for handling the necessary data flow, examining respondent error and reactions to such interview techniques, and studying cooperation rates with a panel arrangement. In a special recruitment of an Hispanic subsample, penetration of the touch-tone capability itself, and opinions of such interview techniques will be researched.

2. ***PEAC (Program Evaluation Analysis Computer).*** A portable, microcomputer-based response device, PEAC was developed jointly by the Ontario Educational Communications Authority and the Children's Television Workshop. Primarily used for formative television research, PEAC collects second-by-second viewer responses to television programs, with viewers registering continuous or discrete reactions on any of 16 response keys with variable labels [11].

Color graphics data reporting techniques present viewer reactions time-locked to the programming which stimulated them. Fine-grain analyses are possible, with visual representations of viewer involvement, comprehension and/or enjoyment directly linked to program dynamics (e.g., specific script lines, appearances of a given personality, scene or format change). Data reporting methods, including superimposition of real-time viewer responses over an entire program or segment, are specially designed for meaningful input to program production decisions.

For formative research purposes, the device is usually used in a group setting, with a group leader overseeing the viewing and simultaneous evaluation of programming material.

With appropriate panel instruction, however, PEAC units can be geographically distributed, with viewers taking units home for a week or so of program evaluation. Respondents would have an electronic qualitative diary, entering program evaluation judgments on this device, instead of on a traditional paper and pencil diary. At the end of the diary period, PEAC devices could be picked up by field interviewers and data from each device aggregated in a central computer for analysis and reporting.

Like the PAL system, a PEAC panel could be more easily and inexpensively rotated than the Percy panel. Unlike PAL, PEAC would not require viewers to tie up home phone lines for qualitative television data entry, so panelists might be amenable to providing more extensive program evaluations gathered throughout the program, rather than simply at its conclusion.[4]

Unlike both the Percy System and PAL, individual response devices could easily be made available to more than one household member at a time. Neither PEAC nor PAL can electronically verify program viewing (or duration of viewing), since both devices operate independently of the television set. For this reason, neither can

simultaneously serve to measure household or individual viewing behavior, as the Percy System does.

D. Methodological Alternatives and Considerations

Methodological alternatives for a qualitative television ratings system are as numerous as unresolved concerns about valid collection, analysis and interpretation of qualitative research data. Table 7 presents an analysis of some existing qualitative ratings systems and design features which can be used to distinguish them.[5]

Design variables which differentiate existing qualitative ratings systems begin with the very nature of the qualitative ratings to be collected (e.g., whether they are overall, one-dimensional judgments of programming effectiveness, or more detailed diagnostic analyses of the sources of viewer response). Additional design variables include: when the qualitative ratings are collected (e.g., whether viewers evaluate programs during or some time after viewing); where they are recorded (e.g., whether viewers provide ratings at home or at a special test site); from whom they are obtained (e.g., from a permanent or rotating panel, from a limited or national sample); the data collection channels used (e.g., personal interview, paper and pencil diary, electronic response device, telephone interview); the frequency with which ratings are ascertained (e.g., on a regularly scheduled basis or as special studies are required), and whether or not the system relates qualitative ratings to actual viewing behavior (traditional program ratings).

Two papers commissioned for the conference (Appendices B and C) raised research design and analysis issues generated by current theory and available research data on viewer program evaluations. These methodological concerns fell into three principal categories: first, the nature of the qualitative ratings data to be collected; second, sampling of viewers and of evaluation events; and third, potential applications of research results.

1. The Nature of the Qualitative Ratings Data to be Collected. Many existing qualitative ratings systems rely on (and sometimes exclusively measure) some indication of program appeal. Participants questioned the authenticity of viewing satisfaction data collected by these systems. For example, how "natural" are such evaluation responses, both in terms of the judgmental tasks imposed on the viewer, and in terms of response mode (e.g., button-pushing, scale ratings)?

Most importantly, does the process of evaluating television programs, as imposed by the various ratings systems, fundamentally alter the television viewing process? Is it more acceptable to watch less satisfactory programs, if one can "ZAP" them (Percy System) or lower their "Q"? Conversely, does a straining for consistency of behavior and attitudes foster positive responses to viewed programs? How might the continual evaluation processes superimposed on viewing by the Percy System, PEAC and ASI alter viewer involvement with the program? Is involvement with a "boring" program modulated by the imposition of a simultaneous evaluation activity? Or does a highly involving program impede the program evaluation process? What is the meaning of response rate in such systems?

Table 7

Some Existing Qualitative Television Ratings Systems

System Descriptors	VOXBOX	TvQ	PTvQ	TQR*	CBS Viewer Session Room	ASI	French Diary	British Diary	Canadian Diary
Ratings Collected									
—during viewing	x				x	x			
—after viewing		x	x	x			x	x	x
Ratings Recorded									
—at home	x	x	x	x			x	x	x
—at special test site					x	x			
Uses									
—electronic recording device	x				x	x			
—paper & pencil measures		x	x	x		x	x	x	x
Uses									
—permanent panel of respondents	x								
—rotating sample		x	x	x	x	x	x	x	x
Panel/Sample is									
—national		x	x	x			x	x	x
—limited	x				x	x			
Provides									
—overall program ratings		x	x				x	x	x
—"diagnostic" ratings	x			x	x	x	x	x	x
Surveys are									
—regularly scheduled	x	x		x*			x	x	x
—special studies	x		x	x	x	x	x	x	x
Viewing behavior									
—also measured	x			x			x		
—not measured		x	x		x	x		x	x

*syndication of TQR is anticipated; system operations are described as currently planned

With qualitative rating panels, it is particularly important to consider possible cumulative effects of program evaluation on program selection and viewing processes. Panel effects problems extend to the very psychology of viewer program evaluation. Are program evaluation panels breeding a special group of viewer-critics whose criteria and standards over time come to misrepresent those of the larger viewing population?

These considerations emphasize the importance of understanding the psychology of viewer attitudes toward television programming, and of deriving minimally obtrusive evaluation processes based on customary viewing and perceptual processes.

Participants also warned that viewing satisfaction is only partially derived from program content. Viewing satisfaction is also influenced by a variety of contextual factors, and by individual differences among viewers. Viewing conditions (e.g., the interpersonal environment, presence or absence of secondary activities, counter-programming, and rather idiosyncratic utilities viewers develop with regard to a program) can heavily influence program evaluations. Accordingly, appeal dimensions of a qualitative ratings system should never be expected to explain or predict totally a viewer's selection of, or satisfaction with a given program.

Participants also cautioned that qualitative program ratings, whether registered during or after program viewing may be more so a *post hoc* explanation of past viewing than powerful predictors of future viewing choices or satisfactions. Given a possible consistency bias, in which viewers say they like the programs they view, it is interesting to question the relationship between amount of viewing of a given program series and qualitative ratings of that series. Does appreciation increase with repeated series viewing?

Further discussion questioned the adequacy of an unidimensional qualitative ratings system. Since most existing one-dimensional ratings use an appeal ("appreciation," "interest," "enjoyment," "favorite") dimension, the discussion tended to focus on the adequacy of a viewing satisfaction measure.

Repeatedly, participants advised that moves toward additional dimensions of viewer response to programming must be negotiated in close consultation with the various decision-makers for whom the qualitative data are intended (e.g., producers, Board members, Fund Director, local station programming executives). With this caveat, participants proceeded to examine the uni-dimensional/multi-dimensional issue.

Various research findings suggest that a multi-dimensional approach is essential if we hope to reflect the complexity of viewers' evaluative responses to television programs. Studies by Landis [8] and Himmelweit *et al.* [6] bear evidence of a discriminating viewership with a readiness and capacity for articulate program criticism. Their research also speaks, as do the French studies (see pages 16-20), of the apparent complexity of viewer program images, with overall evaluations constructed of multiple and independent perceptual dimensions. In fact, Himmelweit *et*

al: found the very complexity of a viewer's response to television programs to be strongly correlated with overall program appeal [6].

Analyses of the British "appreciation" ratings uncovered an apparent program type bias, with larger standard deviations in qualitative ratings between different program types than between individual programs within a given type [14]. It may be that an overall unidimensional rating related to program appeal favors certain program types more than others, and that it would take multiple evaluation criteria to provide viewers with an adequate array of judgment opportunities for various program types.

Perhaps even more important than the scientific evidence in favor of multiple program evaluation criteria is the contention that a one-dimensional qualitative measure could never hope to address the varied programming objectives of public television. Participants argued that the full summative and diagnostic potential of a qualitative ratings system would not be attained unless the system encompassed a variety of cognitive, affective, aesthetic and possibly behavioral viewing impacts.

This argument was extended to the extreme case of the special summative research project where programs are extensively analyzed for their ability to facilitate specific desirable viewing impacts (e.g., information gain, reception and retention of attitudinal statements, additional information sought from non-television sources). Participants acknowledged the exceptionally high cost of such research and the ultimate incomparability of research designs and findings across summative studies of different television programs.

Between the two extremes of a simplistic one-dimensional rating and costly summative studies on individual programs, participants discussed the possibility of an intermediate approach. A two or three-dimensional ongoing ratings system could be designed to tap program appeal and one or more indicators of cognitive or affective response. This overall ratings system, with its more superficial analysis of all programs, could then be supplemented by more extensive summative evaluations of a subset of goal-directed programs.

To identify the two or three dimensions for the overall ratings system, participants discussed the possibility of collapsing the TQR factors, to derive some higher-order solution of the data. Whatever the approach, they emphasized the importance of deriving the ratings criteria from viewer perceptions of PTV, rather than of television in general. While participants agreed it would be useful to compare public and commercial programming along the same dimensions, they felt viewers might have different expectations for public television, and that the rating dimensions ultimately selected should reflect those expectations and related experiences. In this light, it was suggested that the TQR procedure be specially repeated to build an evaluative perceptual space solely on public television programming.

2. Sampling of Viewers and Evaluation Events. Participants discussed sampling considerations from several perspectives.

From the perspective of "people sampling," sampling theory would call for a

random sample of viewers. Since the true distribution of qualitative viewer responses is unknown for the total viewer population, it cannot be argued that a demographically balanced national sample will adequately reflect various programming "taste cultures" and preference groups. Random sampling, the only acceptable approach, is impossible, given the double self-selection process in most qualitative ratings systems. Respondents and panelists must not only self-select into the viewing audience, they must also choose to cooperate with the evaluation procedures.

Sampling theory aside, participants expressed great concern for demographic considerations, insisting that the system not "disenfranchise" significant population subgroups. Analyses of various data collection channels can be compared in terms of cooperation rates obtained among ethnic minorities, the poor, and the elderly.

Another people sampling consideration is the nonviewer. Many of the systems discussed sample current viewers of the programs being evaluated. One innovative alternative called for a "nonviewer coincidental" in which nonviewers of the program studied would be sought out, to study what other program they were watching (if any) and to assess perceptions of the nonselected program.

People sampling considerations include the observation that very different kinds of viewer/critics may be contributing equally weighted (and perhaps highly divergent) qualitative ratings to the same data base. It may be helpful to separate these viewer/critic types in subsequent analyses. Different viewer subgroups, for example, may give more salience to different program evaluation criteria [14]. The relative proportion of these subgroups within the same program evaluation audience could be a very useful figure in interpreting that program's subsequent qualitative ratings. Should a multi-dimensional qualitative ratings system segment viewer preference subgroups, in terms of the evaluation criteria most important to each?

Goodhardt *et al.* [5], for example, found systematic differences between light and heavy viewers of television, with heavy viewers generally offering more positive program evaluations than light viewers. This finding becomes valuable when related to another finding, that programs with large audiences tend to have a higher proportion of light viewers than do programs with smaller audiences. The implication for a program with a large audience is a qualitative rating lowered by the light viewer representation. Should a qualitative ratings system segment viewer groups by amount of television viewing?

In addition to people sampling concerns, participants discussed the sampling of evaluation events over time. An individual viewer may expect different responses from a given program, depending on situational viewing motives and conditions. How much might individual preference judgments be expected to vary across such different viewing intentions and experiences? In this sense, it is important to distinguish between systems which provide episode-specific evaluation data (sometimes time-locked to program viewing), and systems which tap overall program

images. For example, PTvQ and TQR are intended to address broader program perceptions, with the respondent's frame of reference not directed to any one program episode.[6] The Percy System data, on the other hand, might be expected to tap fluctuating viewer responses to programs viewed and evaluated, and to produce a far less stable qualitative rating. Participants recommended repeated measurements over time to examine the extent of such fluctuations and the relative stability of various qualitative variables. Successive studies over time may be necessary for yet another reason. Current program evaluations are, at least in part, a function of existing television fare and industry practices, and of contemporary viewer standards and sensibilities regarding television treatment of various subject matter. As these conditions change, possible related changes in viewer program evaluation criteria should be monitored.

Finally, most existing qualitative ratings systems, even if they do propose to track future viewer and programming trends, are still based in the evaluation of available television fare. They contain the viewer/critic within an assessment of current programming fashions, curren perceived roles for television, and current viewing practices and environments. How might a qualitative research system encourage viewers to tell us what they want television to be, not just how they respond to its current performance?

3. Potential Applications of Research Results. Participants questioned whether the various purposes for which qualitative data might be collected could be successfully addressed by any one ratings system. Formative and summative applications of the data, for example, would dictate entirely different sets of information needs and related data collection requirements.

It was recommended that OCR strongly consider the development of multiple qualitative ratings systems, each with special research purposes and capabilities. Pilot studies, for example, might be the forte of the Percy System, where second-by-second viewer response can be time-locked to specific script lines, program personalities or story events. TvQ might contribute to a better understanding of national awareness problems with PTV programs and related needs for promotional work. TQR might best be employed for an analysis of PTV's institutional image as a programming source, and for in-depth analysis of a given program's positioning within that image. A nationwide PAL panel might serve for overall qualitative "overnights," while qualitative diaries would have a slower turnaround but provide more extensive data. Rather than attempt to impose multiple research application purposes on any one method or system, participants discussed a comprehensive qualitative research program in which a variety of research capabilities would address a range of related information needs.

Finally, if future qualitative research efforts extend to the study of cognitive, affective and behavioral impacts of programs, correct interpretation and application of the data will require an understanding of potential interactions among such viewing responses. Varieties of qualitative response may summate, interact and/or

directly compete with one another. Appropriate use of such research results will have to be guided by relevant theory and research findings. The qualitative research movement suffers, in general, from the lack of a theoretical base with which to identify key qualitative research variables and guide analyses of their interrelationships.

III. RECOMMENDATIONS FOR THE DESIGN OF A QUALITATIVE RATINGS SYSTEM

Participants' recommendations to the Corporation covered four principal areas: first, the positioning of a qualitative ratings system within OCR's total research program; second, the nature of data to be collected by such a system; third, methodological considerations in the collection and analysis of these data; and fourth, approaches to maximizing the applications value of these data to PTV programming decisions.

A. The Larger Research Context

A qualitative ratings system should be embedded within a comprehensive and internally coordinated qualitative research program with a range of research purposes and capabilities.

For example, conference participants urged that a simple overall ratings system (like the Canadian enjoyment index) be supplemented with related but more detailed special studies (e.g., studies of PTV's institutional image or of uses and gratifications associated with PTV, TQR-type studies of the nature of viewer program perceptions). When the simpler, overall ratings system is more affordably implemented on a regular basis, it can be used to build norms and analyze trends over time, and findings can be more simply communicated to data users. Ideally, special studies would cover more basic research questions in much greater detail, generating findings useful in updating and refining the simpler ratings system.

Furthermore, where a variety of qualitative research studies are initiated to address different information needs, it is important to encourage coordination among these several approaches, interrelating findings to build a knowledge base. For example, TQR studies might be used to identify improved button labels for the Percy System, or qualitative profiles of PTV's institutional image might suggest specific evaluation criteria for future pilot testing efforts.

The context of a qualitative ratings system, that is, its position within a larger program of coordinated qualitative research projects, is yet embedded within the total audience research package. The research cycle can parallel the program life cycle, tracking and contributing to: program concept development, concept testing, formative research, pilot testing, traditional program ratings and special impact studies. Additionally, these programming and parallel research activities should exist within a still larger institutional effort to understand how the public perceives PTV, what they would like it to be, and how to minimize the discrepancy between

the two. The result of visualizing qualitative research within these larger contexts is a multi-tiered research program, with coordinated research activities spanning a wide range of highly related programming goals and activities.

B. The Nature of the Data to be Collected

Repeatedly, conference participants emphasized that the qualitative ratings system envisioned by CPB should be tailored to the programming goals and objectives unique to public television. Collectively, qualitative ratings data from both commercial and public television will allow important comparative analyses of public perceptions of the programming source types, including explorations of programming "gaps" in which PTV might provide publicly desirable fare. However, both comparative and separate analyses of PTV programming should be conducted in terms of the specific evaluative dimensions most meaningful to PTV programming decision-makers and evaluators. Accordingly, the various groups of PTV programming decision-makers (CPB Board, Program Fund, program producers, local station programming executives) should be consulted closely throughout the design of the new qualitative research program. Otherwise, the new research data will have no hope of credible reception and application by those parties ultimately responsible for the viewer's experience with public television.

Designing qualitative research efforts for maximum relevance and utility will require careful articulation of objectives by the various programming decision-makers in public television, an articulation which demands close interaction between researchers and decision-makers.

Even in a close working formative team of producers and researchers, such goal-setting is extremely difficult. For PTV, with its decentralized programming arrangements at the local level, and with divided responsibilities for program planning, funding and evaluation at the national level, qualitative researchers will have to foster articulation of programming objectives by multiple "clients" with varying information needs.

Given the multi-tiered, multiple studies approach recommended above, these varying information needs should be addressed by separate, but interrelated research efforts. Wherever possible, the various research methods used should be standardized to permit the accumulation of an integrated knowledge base and a cross-fertilization of methodological refinement. Beyond this core of standardization, however, flexible research procedures should respond to the unique and varying requirements of different types of programming decision-makers.

With great concern that the specific qualitative research variables to be measured be finalized in close consultation with these various "user" groups, participants addressed the issue of whether or not a uni-dimensional measure of program appeal would be an adequate qualitative rating for public television. Generally, there was strong discomfort with a one-dimensional rating, yet deep appreciation of its simple appeal as an "other" number readily juxtaposed with traditional ratings.

Rejecting the uni-dimensional model, participants turned to a two-dimensional

approach, a two-part rating lending itself to simple manipulation and presentation, but venturing into multiple PTV programming values and acknowledging the complexity of the viewer's experience with quality programming. Such a two-dimensional qualitative rating could cover enjoyment (program liking, viewing satisfaction, appreciation, etc.) and a measure of "worthwhileness" (some indication of whether or not the viewer benefited from seeing the program, "got something out of it").

Several participants preferred to include a larger number of rating dimensions, a wider range of qualitative research variables (e.g., a set of TQR dimensions or a list of viewing gratifications obtained by viewers). They argued that multiple dimensions are necessary to capture the full complexity of viewer interactions with programming and that a single qualitative rating index could be derived from them to achieve the elegant simplicity of a uni-dimensional rating, but grounding that one number in extensive backup data which can also serve formative and diagnostic purposes.

A qualitative research program is, by its highly exploratory and rather irreverent nature, one which can break free of the limitations of television as it is, to explore television as it might be. Most existing ratings systems focus on a descriptive evaluation of television as it is, with the research limiting itself to viewer evaluation of current fare, but unable to stimulate either the viewer-respondent or the reader of resulting research reports to normative judgments about the potential of the medium.

To this end, it was recommended that one approach to the design of a qualitative rating system might originate with an ascertainment of public attitudes toward television and its perceived ability to address personal, community and national problems currently unaddressed by television programs. If, for example, PTV program planners at the national level were amenable to such an approach, then future program funding efforts could systematically seek out programming which addressed these social needs. Subsequent qualitative ratings would then evaluate the extent to which viewers believe resulting programs successfully addressed the problem areas.

C. Data Collection and Analysis

While the most critical research design questions focus on identifying those qualitative ratings variables with the greatest decision-making relevance for public television, several data collection and analysis issues will also impact on system design considerations. Small-scale field tests and preliminary research analyses were recommended over the next few years, to address these concerns.

Sampling of viewer evaluations of individual PTV programs will be problematic, given the small audiences for many of these programs. Recruitment of viewers to view and evaluate individual programs might boost sample size to an acceptable level, but resulting data would be subject to the criticism that the recruitment "set" sensitized the viewer and distorted the normal viewing process. A second approach might be to prescreen a pool of television viewing households for past PTV viewing

behavior, thus increasing the probability that a sample drawn from that pool, for example, would be identifiable as "natural" viewers of PTV programs in a qualitative telephone coincidental study.

Generally, participants were concerned about the long-term consequences of membership in a qualitative television ratings panel, and they recommended periodic rotation if qualitative rating panels were to be established.

Sampling concerns also emphasized the importance of designing research methodologies which would not disenfranchise population subgroups traditionally underrepresented in social science research. Comparisons of the overall cost-effectiveness of different data collection procedures should include analyses of cooperation rates obtained with each for minorities, the poor and the elderly.

Finally, qualitative ratings must be analyzed in relation to traditional measures of audience size. The precise relationship appears complicated by program typology, scheduling variables, viewers' program selection practices and possible response biases inherent in qualitative ratings.

It may be, on one hand, that qualitative and traditional program ratings are positively correlated. If this is so, then qualitative research efforts would appear to be duplicative, unless the diagnostic value of resulting data can be applied to the development of more successful PTV programming. If, on the other hand, qualitative ratings show no clear relationship to audience size, one must confront the sharp accusation that qualitative ratings are simply verbal artifact and useless in any practical attempts to improve viewers' actual experiences with PTV programming. Whatever the nature of this relationship, it must be understood in order to rationalize the development of new qualitative research measures, to interpret findings correctly, and to draw meaningful implications for future programming activities.

D. Applications of Research Findings

Participants encouraged application of the data to a number of program development and evaluation activities. In addition to the use of these data as measures of program effectiveness, participants perceived the potential utility of the data:

- in the establishment of programming priorities and derivation of related program concepts;
- as formative input to the development of programs targeted to special populations;
- as a basis for the standardization of pilot testing procedures;
- to guide institutional promotion efforts, as well as those related to individual PTV programs; and
- as input to effective program scheduling and audience-building efforts by local stations.

FOOTNOTES

[1] A complete list of conference participants is provided in Appendix D.

[2] In the Corporation's approach, qualitative ratings do not refer to formative research efforts or pilot tests, in which individual programs are tested prior to any national airing. They also do not refer to specially tailored summative evaluations of purposive programs.

[3] The Audience Research Center operates a similar qualitative radio research program, not presented here because of CPB's preliminary focus on qualitative television research.

[4] As response units are capable of storing a maximum of 500 keyboard responses, evaluation procedures which require higher response rates would necessitate more frequent collection of the PEAC units for central data storage purposes.

[5] Each of the systems described in Table 7 is presented in much greater detail in Appendix C, pp. 2-5.

[6] Actually the frame of reference for viewer evaluations is uncontrolled in two respects. Screening for an undefined program "familiarity" combines responses from viewers who have only heard about the program or saw it long ago, with responses from the weekly loyal viewer of the same program. Even given the possibility of comparable viewing histories, it is impossible to know whether the respondent is evaluating a program on the basis of several seasons of past viewing or on the most recently viewed episode.

REFERENCES

1. Anderson, James A. "Methodological Assessment of Existing Evaluative Ratings Systems." Paper commissioned for the CPB Technical Conference on Qualitative Television Ratings, March 9-12, 1980. Appendix C.

2. Bowman, Gary. "Consumer Choice and Television." *Applied Economics*. 7, pp. 175-184.

3. Dordick, Herbert S. "International State-of-the-Art in Qualitative Ratings: An Overview." Paper commissioned for the CPB Technical Conference on Qualitative Television Ratings, March 9-12, 1980. Appendix A.

4. Durand, Jacques and Jean Arachtingi. "Studying the Quality of Television Pro-

grams: The Experience of the Audience Research Center of the French National Broadcasting System." Paper presented at the CPB Technical Conference on Qualitative Television Ratings, March 9-12, 1980.

5. Goodhardt, G.J., A.S.C. Ehrenberg and M.A. Collins. *The Television Audience: Patterns of Viewing*. Lexington, Mass.: Lexington Books, D.C. Heath and Co., 1975.

6. Himmelweit, Hilde T., B. Swift and M.J. Biberian. "The Audience as Critic: An Approach to the Study of Entertainment." In P. Tannenbaum (Ed.) *Entertainment Functions of Television*. New York: Lawrence Erlbaum Associates, 1980.

7. KCTS-TV Research Department. *Final Report on the Evaluation of "Righteous Apples."* Seattle, Washington: KCTS-TV, 1979.

8. The Landis Group. *Mapping Analysis for the Corporation for Public Broadcasting*. West Palm Beach, Fla.: The Landis Group, October, 1979.

9. Lehmann, Donald R. *Choice Among Similar Alternatives: An Application of a Model of Individual Preference to the Selection of Television Shows by Viewers*. Unpublished Ph.D. dissertation. Purdue University, 1969.

10. Monoghan, Robert R. *Television Preference and Viewing Behavior*. Unpublished Ph.D. dissertation, Michigan State University, 1964.

11. Nickerson, Rand B. "The Application of Microcomputers to the Evaluation of Instructional Media." Paper presented at the Third Canadian Symposium on Instructional Technology. Vancouver, British Columbia: February, 1980.

12. Office of Communication Research, Corporation for Public Broadcasting, *Newsweekly: Audiences, Producers and Researchers Collaborate to Develop a TV News Service*. Washington, D.C.: Corporation for Public Broadcasting, 1979.

13.——— "Preliminary PTvQ Findings." Paper presented at the 1979 NAEB Convention. Chicago, Ill.: November, 1979.

14. Philport, Joseph C. "The Psychology of Viewer Program Evaluation." Paper commissioned for the CPB Technical Conference on Qualitative Television Ratings, March 9-12, 1980. Appendix B.

15. Roll, Charles W. Introductory Remarks, CPB Technical Conference on Qualitative Television Ratings, March 9, 1980.

16. Watt, James H. "Television Form, Content Attributes and Viewer Behavior," In Voigt, Melvin J. and G.J. Hanneman (Eds.) *Progress in Communication Sciences*, Vol. I. Norwood, N.J.: Ablex Publishing Corporation, 1979, pp. 54-90.

APPENDIX A
INTERNATIONAL STATE-OF-THE-ART IN QUALITATIVE RATINGS: AN OVERVIEW

Herbert S. Dordick

Annenberg School of Communications

University of Southern California

TABLE OF CONTENTS

		Page
I.	Introduction	A-1
II.	Sweden	A-2
III.	The Netherlands	A-4
IV.	France	A-5
V.	Great Britain	A-9
VI.	Japan	A-13
VII.	Canada	A-15
Sources		A-18

INTERNATIONAL STATE-OF-THE-ART IN QUALITATIVE RATINGS: AN OVERVIEW*

I. INTRODUCTION

The viewer (or listener) has always had a limited ability to talk back to the receiver. Turning the set off, the most meaningful message to the broadcaster and sponsor, has been most difficult to do. Other alternatives, such as refusing to purchase the advertised products or contributing to the public station has had little or no impact on the program provider. And in most countries in the world the owner of the receiver has no choice but to pay the license fee or tax. Political actions are possible but unrewarding both here and abroad. The fact is, viewers and listeners have a very limited vocabulary with which to communicate with the providers of their radio and television fare.

Qualitative ratings offer prospects for an expanded viewer/listener vocabulary and a feedback channel. Qualitative ratings also offer the decision-makers responsible for broadcast policy increased information concerning the effectiveness of their actions.

The type of rating system designed and its applications depend on the nature of the broadcast system in which it is used. Throughout the world, broadcast systems are being challenged, not only as to their roles in society but as to how they are performing these roles. In part the new broadcast and non-broadcast technologies are opening channels of distribution, altering the shape of broadcasting, and providing opportunities for programming and entrepreneurship. But equally important are the world-wide changes in societal values that are placing new demands on broadcast institutions. Programmers want more editorial and artistic freedom. Viewers want their diverse life-styles reflected by their media.

Both quantitative and qualitative measures of broadcast performance are important to those seeking change. It is abundantly clear that existing measures and methods for obtaining these measures are incomplete, inaccurate and increasingly expensive. But they are also necessary if we are to make better use of the resources for broadcasting. As these resources become scarce in the face of competing demands for energy and other needs, how we obtain these measures of broadcast performance is most important.

---

* This report has been drawn, in part, from "Broadcast Practices in Europe and Canada: An Analysis of Relevance to Public Broadcasting in the United States" prepared by the author for the Carnegie Commission on The Future of Public Broadcasting in August 1978.

Efforts to improve the quality of qualitative evaluations are going on in almost every major country in the world, and especially in those nations where broadcasting has come under the greatest challenge. Thus Britain, France, The Netherlands, Sweden and Canada are all taking steps to increase performance and lower the costs of research. In Japan (reported here), Italy and West Germany (both not reported) there appears to be little activity except in the performance of basic research that seeks to better define the social impact of broadcasting. But because of rapidly changing forces, by the time this report is completed and published, qualitative rating activities in these nations might very well blossom. Indeed, as we go to press, we have been informed of changes in procedures already underway in those nations for which we have provided summaries.

II. SWEDEN

The operation of broadcasting in Sweden is divided between the Swedish Telecommunications Administration which is responsible for the establishment, maintenance, and operation of transmission and distribution facilities, and Sveriges Radio (SR) which has the sole and exclusive right of determining what radio and television programs should be broadcast. The Telecommunications Administration is organized as a central administrative agency and run as a public commercial enterprise. A joint-stock company, SR has the following shareholders:

1. The "popular movements" and coordinating organizations for the various movements holding 60 per cent of the stock;
2. Newspapers and press organizations holding 20 per cent cent of the stock; and
3. Organizations representing the business community holding 20 per cent of the stock.

There are three operating program units: (1) Radio produces programs for nation-wide distribution and has production facilities throughout the country, (2) TV-S produces television programs for nation-wide distribution with facilities in Stockholm and (3) TV-R produces programs for regional as well as nation-wide broadcast with production facilities outside of Stockholm. There are nine regions, each with an Audience Advisory Council for "discussion and exchange of views on programming."

Swedish broadcasting is financed from license fees and there is no advertising. Parliament (Riksdag) makes special grants for overseas and educational services and may fund audience research projects.

The Swedish Broadcasting Commission stresses that high audience ratings are not a decisive measure of quality; "It is more important to increase the satisfaction of each individual viewer or listener than to increase the amount of time people spend watching television or listening to radio."

The Audience and Program Research Department of Sveriges Radio (SR/PUB), organized in 1960, is an autonomous and essentially independent unit within the

General Staff and Services division of SR with the aim of self-financing by providing research services both to SR agencies and external clients. Applied research is ordered and paid for by clients, while basic research is initiated by SR/PUB and financed partly out of public research funds.

Within this single unit formative research, pilot tests and special summative evaluations are performed, as well as daily qualitative research on broadcast programs during and soon after airing.

With the introduction of two television channels the SR/PUB introduced the *5-Minute Technique* of audience measurement. The basic idea of this approach is to measure viewing during each five-minute period of the program, instead of for the total program as had been done previously. Respondents are asked when their viewing begins, stops or switches from one channel to the other.

Approximately 600 telephone interviews are taken each day, each interview comprising three retrospective program days. For each program the interviewer asks: "Did you see anything of the program?" Should the interviewee spontaneously say that the whole program was watched, the interviewee is marked down as a viewer of all five-minute periods of the program. If this is not the case, the interviewer goes on to ask: "Did you see the program from the beginning to the end?" If the answer is affirmative, the coding is the same as above. If the answer is "no," questions are asked to determine which parts of the program have been viewed. With a knowledge of the program length, it is then possible to determine how much respondents watched within five-minute approximations.

When a respondent has been classified as a "viewer," he or she is asked to rate the program on a five-point scale, from one point (very bad) to five points (very good).

Daily tables and diagrams are constructed which show changes in the size of the audiences for programs on both channels. Monthly patterns of viewing can be generated, allowing the comparison of day-by-day viewer patterns with the average for the previous month.

Periodically, demographic data are collected. The program ratings are made available to the program producers and to system executives.

Daily evaluations are but one segment in the wide ranging research undertaken by SR/PUB. These data eventually find their way into extensive basic and applied research undertaken by the Department including:

- psychographic research which seeks to determine various viewer populations based upon several life-style factors;
- formative evaluation based upon data obtained from the Five-Minute Technique and psychographic analyses;
- summative evaluations "ordered" by specific programmers and production units; and
- longer term research to determine "information gaps" in Swedish society in the search to eliminate "information poverty."

Swedish television and radio achieves a high degree of "viewer talk-back" through these many research projects where, it would seem, all of the households in the country will, over a period of a few years, have an opportunity to respond to either telephone interviews or diary "interviews" by mail. (Sweden has a population of about 8.5 million with about 3.5 million households.)

III. THE NETHERLANDS

The complex structure of broadcasting and many other cultural institutions in the Netherlands conforms to a certain pattern which the Dutch call "verzuiling," translated as "pillarisation." A "pillar" usually means a group of organizations or institutions which are based on the same religious or spiritual principles. Beginning with the two major pillars, the Roman Catholics and the Calvinists, the concept has spread to other religious groups, trade unions, professional societies, political parties and with the "Open Door Policy" that emerged in 1969, to almost any nonprofit organization. These "pillars" are given access to both radio and television broadcast time in proportion to their numbers.

Transmission facilities for radio (three channels) and television (two channels) are operated, maintained and developed by a company established by Parliament known as NOZEMA. NOZEMA has no responsibility for programming; they operate in a manner that in the United States would be called a common carrier. There are, however, no rates charged to the users of the transmission systems nor do the users have to concern themselves with the technical problems of transmission and coverage. The numerous programming organizations, representing five major "pillars" and as many as a dozen smaller organizations, can devote their energies to the preparation of programs to fulfill their franchised responsibilities.

Transmission time is allocated in accordance with the number of members in these broadcasting organizations. These organizations publish program magazines designed to attract subscribers who are also counted as members of the broadcast organizations.

The responsibility for broadcasting policy rests with the Minister of Culture, Recreation, and Social Welfare. The Dutch government exerts no censorship. Broadcasting is financed out of license fees but another source of income is advertising. Net proceeds from this advertising are allocated to the broadcasting system as well as a small portion to the press to compensate them for the loss of advertising revenue.

This extraordinary democratic system is matched by an extensive audience research program which seeks to obtain talk-back from all segments of the pluralistic population. KLO (Kijk-en Luisteronderzoek) services all the broadcasting organizations.

A panel of 1500 people, aged 12 and over, renewed annually, keeps a radio diary during four two-week periods in the year, registering each quarter hour spent

listening to radio and which station. The audience data are shown in totals per 15 minutes and in sub-groups according to age, education, profession and other factors.

A different 1500 person television panel keeps a diary throughout the year, processed weekly, to register which programs on each channel are watched for at least half the broadcast. These data provide audience figures for each program and audience flow across various demographic factors.

The KLO uses an evaluation system in these diaries with a grading scale from one to 10 which is available for the total viewing population as well as for the demographic sub-groups. Average evaluation figures as well as degree of variation are reported for each television program.

Set metering instruments are also available to a smaller sample of households. These data simply provide quantitative measures of viewing as well as of channel switching during a program.

On 30 Saturdays a year a sample of 500 people, 15 years old and over are polled by telephone for information on program impact, television viewing motivations, viewers' opinions regarding program and broadcast policies, and the population's attitudes, opinions and knowledge concerning topics treated in the programs. Finally, there are quarterly interviews with the radio panel, ad hoc surveys and screenings followed by discussion or the completion of questionnaires by the panels.

In general, KLO performs most of the qualitative research while a private firm is contracted to perform the quantitative surveys. Each broadcast organization has at least one staff member with the appropriate background for interpreting the findings and deciding what research projects KLO should undertake.

Research findings are made available to program producers and government policy makers. The information is also available to all interested outsiders, such as advertisers, and audiences themselves. When Parliament is considering changes in the license fees this information is widely discussed in the press.

The Netherlands, about half the size of Maine, is a nation of about 14.5 million inhabitants with about 5.5 million households. It is a highly pluralistic society and its broadcast system reflects this pluralism. Television penetration is close to 80 per cent and telephones are in over 70 per cent of all homes. As in Sweden (and even more so) over a very few years, it is highly likely that everyone over the age of 12 will have talked back to their radio or television set.

IV. FRANCE

Broadcasting in France is in turmoil today. There are those who hold that since the end of the Second World War this has been a normal state of affairs. It is often forgotten that prior to the Second World War, broadcasting in France was a rich mixture of both public and private companies. There is growing evidence that this state may return in the near future.

In August of 1974, a number of legislative reforms were introduced in the French broadcasting system. Radio and television channels were directed:

1. to increase program diversity,
2. to separate news and public affairs from entertainment programming and attract qualified professional staffs,
3. to foster competition between the various services or channels,
4. to remove bureaucrats from the creative side of broadcast business, and
5. to tighten control of advertising, imposing significant limitations on time.

Prior to 1974, the production and management of radio and television programs in France was handled by one organization, the Office of Radio-Television Francaise. In 1974, that office was replaced by a complex system of competing authorities:

Three program companies:

— Television Francais I (TFI)

— Antenna 2 (A2)

— France Region (FR)

One radio programming company:

— Radio France

Societe Francais de Production (SFP)—A production company

Telediffusion de France (TDF)—The transmission network

Institut National de L'Audiovisuel (INA)—The audiovisual research organization and library

Policy statements issued by a consulting committee (Le Haute Conseil de l'Audiovisuel) appointed by the Premier set forth very specific programming goals for the program companies as well as for the support groups. These goals are important to the performance evaluations carried out by various research organizations.

License fees and advertising revenue constitute the support for the system. How these monies are distributed is determined by a complex formula depending upon the quality of the programs as well as the size of the audience. Both TFI and A2 compete for advertising revenue, but such revenue cannot exceed 25 per cent of the total budget. (As of 1978, FR3 cannot accept advertising; given the turmoil the system has been experiencing, these conditions may no longer hold.)

The allocation of license fees and advertising revenue is determined by a "Commission de la Qualite" composed of 27 members appointed by the "Haute Conseil de l'Audiovisuel."

The Commission evaluates the quality of a channel's performance along the following lines:

- the programming itself in terms of technical and artistic quality,
- the extent to which original programs were developed, and

• the extent to which programs for the young have been presented.

A public poll is also taken to ascertain the same information.

Daily surveys of almost 1000 persons are taken to determine viewer numbers, as well as attitudes towards the quality of programming. These surveys are commissioned by the Centre d'Etudes d'Opinion and performed by private firms.

A diary is used which covers a seven-week period. The respondent is asked to return the completed sections of the diary before noon on Monday. While the diary is delivered in 14 sections, two for each of the seven weeks, two sections are detached at the end of each viewing/listening week and mailed to the interviewer in pre-addressed stamped envelopes.

A designated or assigned person in the household is asked to record viewing times as well as certain attitudes and other opinions for the entire viewing/listening day. The respondents are asked not to modify their media behavior in terms of amount of viewing or attention to the programs while using the diary. Each week the diarists send in two mailings.

1. The first mailing covers programs viewed Monday through Thursday and evalutes how "interesting" a program is on a 1-6 scale.

2. The second mailing covers programs viewed Friday through Sunday and in addition to the 1-6 overall ratings, several special questions are asked with regard to the entire week's programming and television in general. Table I identifies the special questions which are rotated across the seven-week period covered by each diary.

A few words describing the specific questions asked are in order. Questions about specific radio programs are asked and the listener states the extent of agreement or enjoyment of the program. There are several questions that seek to determine the quality of reception of FM stations and if the listener can receive these programs stereophonically. (Does the listener have the necessary equipment and is the signal well received?)

Another set of questions seeks to determine if the viewer/listener took advantage of the simultaneous radio and television transmission of a program and, further, if the listener used the radio to pick up only the sound portion of the TV program. As in the United States, the French are seeking to improve the quality of sound received without, at least at this time, redesigning the television receivers. Finally, the viewer/listener is asked to rate the importance of this improved sound reception.

Several questionnaires inquire into where television is viewed in the home, at what time of day, whether alone or with children and/or spouse. With the introduction of TV games and teletext, questions are asked concerning attitudes towards these services.

Why the listener listens to radio is probed in another question. For each period of four hours in a day, listeners are asked to state if they listened during this period and why. Did they listen:

— to obtain information
— to keep oneself company

Table I
Qualitative Questions in French Television Diary

Week #1	Week #2	Week #3	Week #4	Week #5	Week #6	Week #7
Identify 2 programs which entertained you, 2 which informed you, and 2 which enriched you. Profile these 6 programs, by indicating whether or not statements apply to them.	"	"	"	"	"	"
Rate each network's programming and provide an open-ended explanation of your ratings.	Indicate your degree of interest in (1-4 scale) various program types and the extent to which current programming in those categories satisfies you (1-2 scale).	Indicate your degree of interest in each of the networks (1-6 scale) and profile each network along 9 descriptors.	Indicate which network you prefer as a source of news and public affairs programming.	What are your motivations for viewing TV? Indicate whether programming from each network is improving, getting worse, staying the same	Special radio questions.	Special radio questions.
Generally describe each network's programming (open-ended).			How do you feel about giving up one night of TV a week? How would it affect your behavior?	What are your PM viewing habits (i.e., who decides to view? What hour do you view? Do you view during dinner?)		

— as a means of accessing cultural programming
— for relaxation

These questions are followed by a set of statements with which the listener is asked to agree or disagree.

These statements allow the viewer to profile a station as one:

— that enables me to pass the time pleasantly
— that permits me to understand what is happening in the world around me
— that tells me what is going on in the community
— that enables me to get through the day
— from which I can receive programs of considerable variety
— that is lively
— that permits me to learn something
— that presents well produced programs
— that permits the expression of all opinions

The two-level evaluations undertaken in France are especially important today as broadcasting is undergoing change. Once again, there are rumors of a major overhaul of the system and there is increased pressure for private ownership of broadcast facilities. Indeed, "pirate" radio stations are already sprouting in the country.

V. GREAT BRITAIN

There is a little need to describe broadcasting in Great Britain. Indeed, there are some who argue that in certain quarters we may be too familiar with British programming. We shall concentrate solely on audience research and in particular on the major interest of this review, on qualitative research.

By law, both the British Broadcasting Corporation and the Independent Broadcasting Authority are charged with the study of reactions, needs and interests of the public they serve. The BBC, under law, are obliged to provide "suitable and sufficient means for the representation to the Corporation of public opinion and the programmes broadcast in the Home Services and for consideration within the Corporation of criticism and suggestions so represented." The IBA are under a similar legal obligation to "ascertain the state of public opinion concerning programmes (including advertisements) broadcast by the Authority."

The Audience Research Department of the BBC states the following objectives for their work:

• Audience Research is the systematic study of the public for broadcasting. Its function is to supply information.

• While it is not by any means the only source of information available to the

Corporation, it is distinctive in that it applies the scientific method, through the use of specialized techniques, to the collection and analysis of information.

- The Audience Research Department serves all departments concerned with domestic sound and television broadcasting. It is not concerned with External Broadcasting.
- The conduct of audience research, and the communication of its results to the rest of the Corporation, constitute the limits of the department's responsibility; its findings are never mandatory.

The kind of information the Department supplies can be summarized as follows:

Background information about the population including receivers in use, people's availability to listen and view, program tastes of the population and their needs from broadcasting, and their interests and spare time activities.

Audience size: estimates of size by sex, age, education, "class," audience flow and overlap, including estimates of program/channel loyalty and patterns of listening and viewing.

Formative evaluations including studies of projected audience as well as pretesting of pilot programs and content analyses of programs.

General Audience studies which are, essentially, summative evaluations of program services as a whole or by particular programs or series, with special emphasis on audience reactions.

Particular Audience Studies including audience reactions, gratifications derived from the listening or viewing experience, short term and long term effects of the experience.

Promotion, i.e., the use of published sources of program information and "training."

The two most performed and routine tasks the department undertakes are: first, to estimate the size of the audience on BBC national television and radio services and on Independent Television and second, to assess the reactions of the program audience, i.e., "the extent to which people enjoy the programmes they see or hear."

Audience Research for the BBC has developed a procedure of questioning samples of audiences after they have listened to or viewed programs. The results are presented in the form of Reaction Profiles which show how those who chose to view or listen to a program "rated" it in a number of different respects. In many cases, a report on the nature of the audience reaction is also provided.

A panel method is used to collect the data. There is a Viewing Panel of between 2000 and 2500 individuals and a General Listening Panel of about 4000 members. Panel members are recruited either by public appeal or by invitation to persons who have been previously interviewed by the Daily Survey interviewers. Since viewing appeals usually bring more applicants than there are places on the panels, it is possible to select members in a manner that makes the panel broadly representative of the population for each service. This can also be done by means of direct invitation by the Daily Survey interviewers. Panel members serve for 12 months.

Each panel member is sent weekly bundles of questionnaires about a wide variety of programs. The panel members are expected to answer questions only about those programs listened to or viewed and, further, not to alter their media habits. This rule of maintaining their listening or viewing habits is stressed over and over and it appears to be working, if one judges by the varying number of questionnaires returned. The larger the program's audience, the higher the proportion of the panel which returns the research questionnaires.

There are several types of questionnaires. Elaborate questionnaires ask about specific facets of a particular broadcast. Questions about a play will cover acting, production quality and script; questions about a variety show will inquire about the viewer/listener's opinions about each act. Usually, but not always, each question is followed by alternative answers from which panel members select that which most nearly corresponds to their feelings about specific facets of the program.

For each program researched, Reaction Profiles and Indices are developed. Each scale consists of a pair of "opposite" adjectives or short descriptive phrases, for example "interesting/boring," "funny/unfunny" with five scale points between them. The panel member is told to indicate his/her rating of the program by selecting one of the five spaces in accordance with his/her feelings about that aspect of the program being studied. An overall rating scale is provided which serves as a general evaluation of the program. The overall rating might be "interesting/boring," "entertaining/dull" or "important/unimportant." The lowest number of rating scales used is four and the highest is 12.

Reaction Profiles show the ratio between the extremes and are often provided in graphic form. For example a program receives a Reaction Profile index of "interesting/boring" (80:20), "well presented/poorly presented" (55:45), "funny/unfunny" (45:55) and so on. Standard sets of rating scales have been developed for each program category, but there is a provision for the addition of one special scale for programs with unusual characteristics. Much importance is attached to having a standard set of scales so that the profile for any given program can be compared with the current average for that genre of programming.

A Reaction Index for each researched program is publishd in the monthly BBC Audience Research Newsletter. This Index is derived from the numerical rating scales. A+ is equal to 4, A to 3, B to 2, C to 1 and C- to 0. The total marks a program receives from the Panel are expressed as a percentage of the maximum possible and defined as the Reaction Index.

All panel questionnaires provide for viewer/listener comments which are usually quite faithfully reported in the Newsletter. If a series is being researched the findings are compared to previous weeks' results.

Each questionnaire requests in bold print at the top: *No Duty Viewing Please*, or *No Duty Listening Please*. The BBC Research Department makes it quite clear that Reaction Indices and their Profiles are no more than convenient summaries. "The answers to other questions and the often voluminous comment made by panel

members provide the real meat." The Department also states that there can be no greater mistake than to assume that an apparently favorable Profile necessarily denotes an artistically or esthetically good broadcast, or an unfavorable Profile a bad one. The public's judgment may not "agree with informed professional judgment." Profiles are not absolute measurements; their primary value lies in facilitating comparisons between the reactions of the audiences of different broadcasts and at different times, but within the same program genre. Ideally, every Profile should be supported by detailed results of the panel's comments, but this is much too costly and time consuming. Some 15 television programs and 25 radio programs are researched by their respective panels each week.

For ITV, measuring the audience is even more important; their revenue depends on being able to assure advertisers that their programs are being watched. The Joint Industry Committee for Television Advertising Research (JICTAR) was organized in 1961 and represents three organizations: the Independent Television Companies Association, Incorporated, the Society of British Advertisers, and the Institute of Practitioners in Advertising. It employs an independent research agency (in 1977, the Audits of Great Britain) to measure the size of the audience. For each ITV area, a long-standing panel of households is selected to represent households in the area that have sets capable of receiving all television programs. These sets are fitted with meters which record when the set is switched on, the channel selected and when the set is switched off. To check on these measures, members of the households periodically log their viewing in individual diaries.

In the 1974-75 period the cost to BBC of their daily surveys was about £450,000 or 62 per cent of all the money spent on audience research. The JICTAR service during that same period cost about £550,000.

Both BBC and the JICTAR are concerned about the rising cost of audience measurement. Further, the discrepancies between the ratings produced by the two systems could result in both systems becoming discredited in the eyes of the press and the public. Both IBA and BBC recognize that there is no relation between the size of an audience and its appreciation of a program. The IBA has undertaken continuous audience appreciation surveys based upon a long standing panel of London viewers which is consulted every two weeks. Additionally there are panels in each ITV region which are consulted twice a year. IBA's Appreciation Indices required an expenditure of £33,000 in 1974-75.

Both BBC and IBA undertake a variety of special studies surrounding special programs and coverage of events such as the World Cup. But the Annan Commission, a blue-ribbon committee established to study the past performance and future potential of British broadcasting, expressed some misgivings about the general quality and value of audience research in Great Britain. Given the cost of the programs and their distribution, the Commission felt that the research efforts have been "too piecemeal, too narrow, and too superficial." Furthermore, the quality of the special studies varied widely; some were quite excellent but too many were not

carefully done and analyzed. While the IBA gathers much less data, they are able to stand back and with the assistance of academics draw broad conclusions about viewers and their attitudes about the broadcasts. In contrast, the BBC is overwhelmed with data and unable to do a credible analysis of long term effects, charting significant trends. There was no coherent research strategy, nor at the time of the Annan Report, any sign of such a strategy emerging. Many producers did not find the work of the research department of much help, and some producers found the research department did not have time to do the studies they requested for some of their programs.

The Commission felt that BBC and the JICTAR should devise a combined system of audience measurement for all broadcast channels, flexible enough to account for new services that might emerge. Furthermore, an independent body should be responsible for the service. The commission recommended that this research be capable of reflecting responses of thousands of viewers rather than millions, to be more responsive to small changes in attendance patterns and the attitudinal factors accompanying them.

There is a barrier of "considerable suspicion and some ignorance" between audience researchers and those doing program production. Similarly, executives of the broadcast organizations also feel somewhat left out of the research process. This, too, leads to suspicion and eventually, both executives and program producers do not effectively use the research results. The audience researchers need to have more contact with both producers and executives in the broadcast departments. The researchers should educate the producers and executives on how the research is being carried out and just how the interpretations have been derived. It is interesting that the Commission noted that the Swedes do a better job of using audience research innovatively and imaginatively.

Basic research on issues such as the place of broadcasting in society, the psychological motivations underlying people's viewing and listening and the economics of broadcasting should be left to academia. But both IBA and BBC audience researchers must keep in touch with this research. But applied research that seeks to serve the broadcasters and program producers should be performed within the BBC and IBA organizations or their assignees.

Audience research can be an aid for the producer and a guide to the schedule; but it cannot take the place of the judgment of eitner. David Hume, in 1751, wrote that the social sciences cannot usurp the role of moral judgment. Inferences from the one are no substitute for the exercise of the other.

VI. JAPAN

Japan's broadcasting services are operated by Nippon Hoso Kyokai (NHK) a special public broadcasting corporation, and by private broadcasters. It is, therefore,

a dual system comprising both public non-commercial and commercial broadcasting. The government's NHK serves the entire nation; while the commercial broadcasters serve their respective local communities. NHK maintains five national networks; two television networks, one each for education and general television; two AM radio networks, one each for education and general broadcasting; and one FM radio network primarily for classical and popular music.

There are 107 private domestic broadcasting corporations which cover 86 per cent of the population as follows:

— 36 Television and AM radio stations
— 55 Television only stations
— 2 AM Radio only stations
— 4 FM Radio only stations

Taking into account the relative size of their economies, Japan's commercial broadcasting is today equal to that of the United States. In major cities NHK face strong competition from at least four and often five commercial rivals. While the 1950 Broadcast Law prohibits commercial stations from owning or operating more than one radio and/or television station, many commercial stations are affiliated with networks—a system that allows them to pool news programs, save on production costs and reap the benefits of national advertising. There are also close relationships between broadcasters and newspapers.

The NHK receives no direct government investment (it proposes the amount of a receiver's fee and the National Diet approves or modifies their proposal) but collects the fee itself, an important factor in their audience research services.

The Public Opinion Research Institute conducts surveys to ascertain: listeners' expectations of NHK programs; audience ratings for NHK programs; public opinion about current issues and about political candidates during elections and time budget studies. The last is an investigation of how people spend the 24-hour day in Japan.

Three per cent (or about $27 million) is spent by NHK on research, a significant portion of which is devoted to an extensive survey system. Additional informal surveys are conducted by the fee collectors taking advantage of their door-to-door contact with viewers and listeners. Program preferences and attitudes are sought. The information is compiled by the Audience Services Center and is provided as basic reference material for overall program management.

Twice a year, surveys are conducted with a sample of nearly 5500 people using the personal diary approach. Listeners and viewers are asked to record their patterns over a seven-day period. Likert scales are utilized for a given genre of programming. For example, the principal factors used in rating drama programs attempt to measure audience responses in terms of: emotional evaluation, dynamic force, outstanding features or characteristics, program structure or organization, and suitability for home viewing/listening. Adhering closely to a uses and gratifications approach, other data seek to determine the degree of satisfaction with specific media for "acquiring culture." These surveys are not limited to radio and television but often

take into consideration newspapers, magazines (both weekly and monthly), books, groups, schools and "specific individuals."

Commercial stations utilize both Japan Nielsen and Video Research.

Observers of the broadcast scene in Japan criticize both NHK and the commercial broadcasters for their heavy dependency on quantitative ratings. Indeed, this dependence seems to be as heavy, if not more so, and generally with less questioning in Japan than in the U.S. In response to this criticism most, if not all, of the broadcasting stations are beginning to develop the "monitor" system.

Every month, a station asks about a dozen citizens with various backgrounds to contribute essays on that station's programs. The essays are issued as a pamphlet for distribution to the managerial and production staffs of the station. An individual can apply to serve as a monitor and a modest honorarium is paid for the essays. These monitors serve semi-annually or annually at most.

While both NHK and various universities have published widely on many of the more serious aspects of broadcasting, there appears to be little serious use of qualitative ratings today. Japan is making much about the emerging Information Society and is documenting in considerable detail how people use media of all kinds. But, to date, almost all of it has been quantitative.

VII. CANADA

Canada has a dual broadcasting system. The Broadcast Act declares that "broadcasting undertakings in Canada...constitute a single system...comprising public and private elements." The public element is the Canadian Broadcasting Corporation (CBC) and is a crown corporation. It operates four radio networks and two television networks, provides a "national service" in French and English. It also operates multi-lingual Northern Services in Indian and Eskimo languages, the Armed Forces Service and Radio-Canada International.

The private sector networks and their audiences are:

CTV Television Network	A national English language TV network
Les Tele-Diffuseurs Associes	French language network of five stations serving areas in Quebec
Global Communications, Ltd.	Regional English language in the Ontario area
Ontario Educational Communications Authority	Educational services in French and English to that province
Office de Radio-Telediffusion du Quebec	French language educational services for Montreal and Quebec
Alberta Educational Communications Corporations	Local radio services for Edmonton and Educational Television in the provinces

In addition there are four independent television stations unaffiliated with any network and purchasing much of their programming from the United States. Finally, there are over 400 independent cable television systems operating in Canada providing both Canadian as well as U.S. programming for over half the homes in the nation.

The CBC is funded from government grants voted annually by Parliament. Commercial advertising sales provide supplementary revenue to the CBC, somewhat more than 20 per cent of the total operating expense. Although CBC radio abandoned commercial advertising in 1975, both English and French television services engage in vigorous sales of commercial time.

In recent years the Canadian government has begun to take a critical attitude toward both television and radio programming, perhaps stemming from a nationwide survey undertaken by the CBC research department in 1974. This research showed that while more than half of the responses stressed the importance of television in the information and news areas, about a third thought radio and television should perform a watch-dog function in public affairs. Large minorities of the respondents felt that television was becoming a "time-wasting" trap and were concerned about the "bad special effects" that might result if television continued to show violence and relaxed attitudes about sexual conduct. Fifteen per cent were hostile towards commercials and, in general, viewers expressed unfavorable attitudes toward them.

When comparing CBC with the private networks, the public seems to feel that CBC is more educational, more informative and "more Canadian." Further, the general feeling was that CBC is a "good thing" and not elitist in outlook. The commercial stations were viewed as "more relaxing, more popular and local." Among both English and French speaking Canadians, American programs were extremely popular. They felt that U.S. programs were more entertaining, better produced, better acted and more varied, albeit not as informative and realistic as the Canadian programs.

Quantitative ratings of Canadian television are obtained, generally, via Nielsen surveys. There is also considerable qualitative evaluation performed in the home via personal diaries. Selected households keep daily diaries recording the programs watched and how much of the program was watched. They are also asked to rate each viewed program on a Likert scale from one (liked very much) to five (not at all). There is also a box to be marked, at the bottom of the left page in a booklet, if none of the programs had been watched for that day.

On the opposite page, a variety of attitudinal questions are asked, including, for example:

— Did you plan in advance to watch tonight's program or did you choose it when it came on?

— How did you find out about the program you watched?

— What were you watching on TV before the *CBC National News* came on?

Additional questions probe viewers' attitudes about the National News program.

For example, here are some comments viewers are asked to consider:

— It was very exciting.

— You had to concentrate on it.

— It was a program you could relax with.

— It was a program you got involved with.

— It made you think.

— It was good company.

— It kept you in touch with what's going on.

Other questions focus on some specific programs, such as series programs. For example, viewers are asked how often they watched the series, how many past episodes they viewed and why they did not view missed episodes (not at home, at home but not watching TV, didn't care for the guests on that particular evening).

Viewers are also asked questions that relate their life-style attitudes to a program episode and its plot. For example:

— Does the fact that Larry King has a new job make tonight's show less or more enjoyable to watch than last season's?

— For the series as a whole, what kind of relationship with the opposite sex would you prefer that Larry King have?

Inquiries are made that seek to determine viewers' attitudes towards specific characters, the story, the setting, camera work and acting.

The CBC research department continues to express interest in viewer perceptions of the program/network or service as; elitist, educational, informative, Canadian, relaxing, popular and/or local.

SOURCES

Annan Report of the Committee on the Future of Broadcasting; March, 1977, London, Her Majesty's Stationery Office

BBC Audience Research in the United Kingdom Methods and Services, 1975, British Broadcasting Corporation, London, UK, 1975

Bill C-24, An Act Respecting Telecommunications in Canada, 1978, Ottawa, The Minister of Communications

Cahiers des Charges, Societes nationales de television et de radio-diffusion, establissement public de diffusion, institut de l'audiovisuel, Paris, Premier Ministre, Secretariat d'etat (porte-parole du Gouvernment)

Geller, Judith F.; *Japanese Public Broadcasting: A Promise Fulfilled*: Aspen Institute for Humanistic Studies Program on Communications and Society, New York, NY, 1979

Grandi, Roberto and Guiseppi Richeri; *Le Televisioni in Europa*; 1976, Milano, Feltrinelli Economica SpA

Hallman, E.S.; *Broadcasting in Canada*: 1977, London Routledge and Kegan Paul Ltd.

Haute Conseil de L'Audiovisuel, *Rapport sur la mise en oeuvre des cahiers des charges des societes de radio et de television*; 1976, 1977, Paris, La Documentation Francaise

Instant World: A Report on Telecommunications in Canada; 1971, Ottawa, Information Canada

Issues in Communications, 1977, International Institute of Communications, London, UK, 1977

Kato, Hidetoshi; *Communications Policies in Japan*, UNESCO, Paris, France, 1978

Mahle, Walter A.; *Communication Policies in the Federal Republic of Germany*, Paris; 1974, The UNESCO Press

Mass Media in the Federal Republic of Germany: Bonn-Bad Godesburg; 1974, Internationes

NOS Informative Radio and Television in the Netherlands; (1978) NOS Information Service, Hilversum, Holland

Pelletier, Gerard; *Proposals for a Communications Policy for Canada; A Position Paper of the Government of Canada*, 1973, Ottawa, Information Canada

Ploman, Edward W.; *Broadcasting in Sweden*; 1976, London, Routledge & Kegan Paul Ltd., London

Radio and TV, 1978-1985: Proposals of the Swedish Broadcasting Commission; 1977, Stockholm, The Ministry of Education and Cultural Affairs

Schramm, Wilbur; *Big Media, Little Media; Tools and Technologies for Instruction*, Sage Publications, 1977, Beverly Hills, California

Van Der Haak, Kees; *Broadcasting in the Netherlands; 1977*, Routledge & Kegan Paul, Ltd.

Williams, Arthur; *Broadcasting and Democracy in West Germany*; 1976, London, Bradford University Press in association with Crosby Lockwood Staples

World Communications: A 200 Country Survey of Press, Radio, Television, Film; 1975, Paris, The UNESCO Press

Yoshida, Jun; *Trends in Television: Audience Survey in Japan; Studies of Broadcasting, March, 1974*, Cultural Research Institute of NHK, Tokyo, Japan 1974

APPENDIX B
THE PSYCHOLOGY OF VIEWER PROGRAM EVALUATION

Joseph C. Philport

The Arbitron Company

TABLE OF CONTENTS

	Page
Introduction	B-1
I. Economic Models of Viewer Program Evaluation	B-2
A. Passive Viewer Model	B-2
B. Steiner's First-Choice Only Model	B-2
C. Wiles' Model	B-2
D. Owen, Beebe and Manning	B-2
E. Bowman's Model	B-3
II. Psychological Models of Viewer Attitudes	B-4
A. A Functional Approach to Measuring Attitudes	B-5
B. Three Conceptual Models of Audience Attitudes	B-7
C. Uni-dimensional vs. Multi-dimensional Approaches	B-7
III. Multi-dimensional Models of Program Evaluation	B-9
A. Himmelweit, Swift and Biberian	B-9
B. Landis Model	B-10
IV. The Program Evaluation Process	B-13
V. Concluding Comments	B-15
Sources	B-16

INTRODUCTION

"TV is a convenience. You don't have to leave the house to be entertained. The great thing about television is that it is there."

"TV is the greatest excuse for laziness ever invented."

"TV can be educational. It can teach you things."

"It takes my mind off troubles. When I watch the soap operas, I realize I don't have it so bad."

"It's a friend, a much needed diversion."

"TV makes you discuss things, things you wouldn't talk about. Like with your children. There was a program called *James At Fifteen* about a boy's first sexual experience. My son saw it, so we had a talk about what the program said. We might not have got to the subject so easily if it hadn't been for that program."

These comments were offered by participants in recent focus group sessions conducted for the Corporation for Public Broadcasting (1978). For whatever reasons, no other medium has ever been discussed, praised, criticized and in some cases condemned as frequently as television. The task of functionally defining audience attitudes toward television programs is not a simple matter. Its difficulty is confounded by not only the range of attitudes toward television but also the variety of ways people use the medium. Steiner's classic study (1963) of the television audience documented television's important role in the American household. It also illustrated that many viewers have mixed reactions to television—seeing both positive and negative sides of the medium, its content, and the way they use it. All media can serve as agents of diversion, entertainment and information. Whereas books, movies, newspapers and radio tend to fulfill specific functions, audiences apply a wide range of functions to television (Katz, Gurevitch and Haas, 1973). Television must address the most diversified set of functions of any medium.

Since the advent of television, most social scientists have focused on the effects of television and more recently the motivations and gratifications that help to explain television (non)exposure. Most researchers and broadcasters have assumed that an individual's viewing of a program is equivalent to a vote for liking or being satisfied with the program. However, research examining the validity of qualititative ratings has found that there is only a minimal relationship between the liking and subsequent viewing of a program (Barwise, Ehrenberg and Goodhardt, 1979). Several television economists have advanced models of viewer program choice that partially explain the weak correlation between liking and viewing a program. Their theories deserve mention in that they consider all components of the media-audience rela-

tionship (including considerations given to the structure of the broadcasting industry, program production and scheduling economics; and audience satisfaction with available programming).

I. ECONOMIC MODELS OF VIEWER PROGRAM EVALUATION

Although most economic models fail to provide in-depth operational definitions of satisfaction, or other viewer attitudes, they do provide a context for critically assessing viewer evaluations that will be defined later in this paper. Five economic models developed to explain the selection of television programs are presented in this section of the paper.

A. Passive Viewer Model (Noll, 1973)

The basic assumption of this model is that individuals decide to watch television rather than specific programs. The true choice is between viewing and not viewing. Given the decision to view, viewers will watch anything rather than not watch television. Viewers will select among available programs but their choice excludes a nonviewing option.

B. Steiner's First-Choice-Only Model (1952)

The First-Choice-Only model is in direct opposition to the Passive Viewer Model. This model assumes that the potential audience can be divided into subgroups with internally homogenous program preferences and that potential programs can be classified in terms of program types. A viewer's preference for a program type is assumed to be exclusive. If a viewer's preferred program choice is not available, nonviewing is always the second choice.

C. Wiles' Model (1963)

This model can be seen as a hybrid of the Passive Viewer Model and First-Choice-Only Model. Wiles agrees with the Steiner model in that viewers have highly divergent preferred types of programs. Considering this divergence, only a small group of the potential audience will be able to select their first choice program type from available program offerings. However, nonviewing is not the second choice. In this model, the second choice is a type of program with some undifferentiated mass-audience appeal. All members of the potential audience will resort to viewing mass appeal programming before resorting to nonviewing.

D. Owen, Beebe and Manning (1974)

These researchers contend that the first-choice-only model should include possible substitutions of a less preferred second or third choice program type for the first-choice program type. Under this model, data on audience size provide inadequate information about viewer satisfaction. Audience size only reveals the extent to which viewers are persuaded to view. As a hybrid of the previous three models, there are three viewing outcomes in this model:

- Viewers will select their unique first choice or select nonviewing as their second option.
- Other viewers will select their closely related second or third choice if their first choice is not available.
- And, other viewers will select their first and second choice, but will watch a mass-appeal program rather than elect non-viewing.

E. Bowman's Model (1975)

Bowman's model assumes that a viewer has an ideal set of characteristics for a program or set of programs. When combined, these characteristics constitute an ideal program that would satisfy the viewer. Faced with not finding his ideal program available, a viewer attempts to maximize satisfaction by selecting the available program that is closest to his ideal. In this model, preference structures are not forced on individuals, but are derived from their own viewing experiences.

These models demonstrate attempts at explaining the complex relationship between television and its audience. Their assumptions generate several questions that should be addressed in a functional review of the process of the audience's evaluation of television programs.

- To what extent is television program exposure motivated as opposed to being almost random?
- Can a single measure of program liking or appreciation adequately or validly measure an audience's reaction to a program?
- Can the audience serve as critic and evaluate stylistic components of program content?
- When should program evaluations be assessed? Are audience predispositions more meaningful than evaluations captured after exposure to actual program episodes?
- Are any types of attitudes better discriminators of programs within one format compared to others?
- What is the relationship between motives for viewing and subsequent program evaluations?
- What types of audience segmentation should be employed to more precisely profile attitudes toward programs?

The recent interest in qualitative programming research has paralleled research efforts that have: (1) systematically tested motives for television exposure, (2) tested the validity of existing qualitative ratings systems, (3) examined new dimensions of program evaluation, and (4) employed multivariate analyses for uncovering the complex relationships between attitudes and viewing behavior. All of these efforts are indicative of the building of a theoretical base that is essential if a functional model for assessing audience attitudes is to be developed. Considering the current state of this research, it is unlikely that a cookbook listing of attitudes would be as

beneficial as a review of some of the major issues relating to the creation of those attitudes and their relationship to television viewing.

II. PSYCHOLOGICAL MODELS OF VIEWER ATTITUDES

To paraphrase Rosenberg (1956) in the context of hypothesizing a person's attitude toward television programs: A person's attitude toward television programs is "accompanied by a cognitive structure made up of beliefs about the potentialities of that program for attaining or blocking the realization of value states." Or, paraphrasing the theory of Fishbein (1965), the following might be applied to a hypothesis concerning attitudes toward television programs:

1. An individual holds many different beliefs about television programs, that is, many different characteristics, attributes, values, goals and concepts that are positively or negatively associated with the programs.
2. Associated with each of these "related programs" is a mediating evaluative response—an attitude.
3. These evaluative responses summate.
4. Through the mediation process, the summated evaluative response is associated with the programs.
5. On future occasions, the programs will elicit this summated evaluative response—this attitude.

Algebraically, the model can be expressed as follows:

$$Ap = \sum_{i=1}^{N} ViIi$$

Where: Ap = Attitude toward TV programs

Vi = value or belief i's importance to the person as a source of gratification, satisfaction or goodness.

Ii = The perceived instrumentality or extent to which the person believes that Programs would lead to or block the attainment of value i or is related to belief i.

In his review of the Rosenburg and Fishbein models, Neilsen (1974) recommends that two unique properties of the Fishbein model be incorporated in an attitudinal model for television programs. Neilsen emphasizes Fishbein's concern that different values and beliefs should be considered for each unique television program. He also contends that while psychological traits may partially explain behavior, social situations or sanctions should also be included in the model.

The new model can be expressed as follows:

$$Ap = \sum_{i=1}^{N} ViIiSi$$

Where: Si = Si is the level of social sanction for satisfying value i through viewing program p.

Neilsen states that one of the truisms of social psychology is that attitudes are learned through social experiences. Attitudes are influenced by other people—the family group, friends, workgroup, neighbors, etc. Unfortunately, Neilsen's study focuses on the evaluation of messages within programs rather than examining the model's application for evaluating reactions to programs.

The use of either of the attitude models discussed above has greater relevance if viewers are predisposed to select specific programs based on their characteristics. Under the Passive Viewer Model, Noll (1973) assumes that individuals decide to watch television rather than specific programs. If this assumption is true, there would be no utility to program specific attitudinal research. Jeffres (1975) has classified the decision to select television as opposed to selecting programs as media seeking behavior and the decision to view a specific television program as content seeking behavior. He concludes that almost three-quarters of television viewing is content seeking. In other words, the Passive Viewer Model only accounts for approximately 25 per cent of the television program selection process. However, it is also important to note that Jeffres makes no distinction between the predisposition to select a program based on its format or its own unique characteristics. We cannot draw any indications as to the appropriateness of the Steiner or Owen, Manning and Beebe models of program choice.

A. A Functional Approach to Measuring Attitudes

Katz (1960) has recognized that some members of the audience may like or dislike a program with equal intensity, but for completely different reasons. The functional approach recognizes that varied motivational patterns in audiences will lead to different attitudes within the audience. Katz contends that attitudes can be classified as having knowledge, utilitarian, ego-defensive or value expressive functions. For example:

- *Knowledge Function*—Attitudes serving this function are closely related to the need to know and understand one's own environment. They relate to the need to process information to give structure to a chaotic world.
- *Utilitarian Function*—Attitudes serving this function are positive towards the means of reaching a desired goal. Negative attitudes are held toward objects or behaviors that are unpleasant.

- *Ego-Defensive Function*—Attitudes serving this function are held to protect oneself from external anxiety or of dangers from the external environment.
- *Value Expressive Function*—Attitudes serving this function help to give the individual positive expression to his central values and self-concept. These attitudes help to confirm one's self-identity.

Gutman (1978) has employed the functional model to segment television audiences. Following the "uses and gratifications" approach (Katz, Blumler and Gurevitch, 1973) and utilizing the Q methodology, his research uncovered six types of viewers based on their motives for television exposure:

- *Dissemblers*—Watch television for in depth news coverage, keeping up with local and national events, improving their minds, getting new ideas and watching plays. However, the programs they viewed covered a wide variety of available programs.
- *Acceptors*—State that they accept television and watch it for a variety of purposes. They also state that television is an acceptable pursuit when there is nothing else to do. This group does not particularly see television as a means for improving one's mind.
- *Information Processors*—This group is characterized as seeing television as a means for keeping informed about new ideas and learning things to improve one's mind. While viewers in this group appreciate in-depth coverage, they seem to be aware of the medium's limitation in its ability to present large amounts of information.
- *Likers* —This group genuinely likes TV, TV personalities and TV series. This group is assertive in its contention that they "do not watch television because it requires so little effort.
- *Thought Avoiders*—This group is interested in news, information and sports. It is not interested in using the medium to follow cultural pursuits.
- *Reality Seekers*—This group watches television to "give me something to talk about with my friend," and watches TV when there is nothing better to do. They view TV for new ideas and to improve one's mind. They tend to watch to pass an hour or because others are watching and tend to have a passive involvement with television.

Gutman's identification of viewer types adds an important consideration to defining viewer attitudes. It is likely that Dissemblers and Information Processors may like the same program for different reasons. Should an attitude scale identify that both groups like the program or give an indication of what components of the program each of the groups like? A second and more essential question is related to the amount of television viewing in each group. Consider a situation where Reality Seekers and Likers are asked to evaluate the same program, and Likers are heavier viewers of the program. The aggregate program evaluation will favor the attitudes of the Likers. While this may appear to be a democratic approach, it may fail to identify

a strong degree of satisfaction in one portion of the audience and a strong degree of dissatisfaction in another portion of the audience.

Having reviewed the basic structural components of attitudes and recognizing how attitudes can vary in specific sub-audiences defined by their motivations for viewing, several models for assessing audience attitudes toward programming can be assessed.

B. Three Conceptual Models of Audience Attitudes

In reviewing literature in social psychology, one finds three major approaches to attitude measurement (Berkowitz, 1972). Each suggests a unique approach for measuring reactions to television programs.

Model 1—Evaluative: In this model, an attitude is conceptualized as a feeling or evaluation. Therefore, an individual's attitude toward a television program would be viewed as a feeling of favorableness or unfavorableness.

Model 2—Conative: Under this model an attitude is conceptualized as a readiness to respond to a television program in a particular way. This readiness could be operationalized as planning to view (or avoid) a future broadcast of a program.

Model 3—Multiattribute: In this model, an individual's attitude toward a television program consists of an interrelated set of understandings, feelings and actions. Specifically, researchers have identified the three components of this model as:

(1) cognitive component—including beliefs and evaluations,

(2) affective component—including feelings and emotions, and

(3) conative component—the behavioral readiness to act.

Tests in both social psychology (Wicker, 1969) and marketing (Bagozzi, Tybout, Craig and Sternthal, 1979) have failed to support the power of single component evaluative models in predicting behavior. Single component evaluative models have typically explained a small but statistically significant relationship with behavior. These finds are supportive of the minimal relationship between viewer attitudes captured in the single component Appreciation Index and viewing of the program (Barwise, Ehrenberg and Goodhardt, 1979).

C. Uni-dimensional vs. Multi-dimensional Approaches

While multiattribute theorists argue that the uni-dimensional approach should be abandoned because the evaluative component, in isolation, is not sufficient to measure attitudes towards television programs, some uni-dimensional theorists argue that the evaluative dimension provides a sufficient definition of attitudes even though the evaluative component may not be uni-dimensional. The multi-dimensionality of the evaluative component of program evaluations was supported in a pretest of a qualitative ratings instrument for the National Citizens Committee for Broadcasting (Philport and DeVerneil, 1978). Viewers were asked to evaluate television programs they viewed during a one week period, using several scales from the

evaluative dimension of the semantic differential (Osgood, Suci and Tannenbaum, 1957). The results indicated that viewers were evaluating programs along two dimensions. Each program was being evaluated with respect to the perceived *Enjoyment* derived from the program and the perceived *Meaningfulness* (or importance or relevance) of the program.

Further analyses indicated that many programs were evaluated as very high in *Enjoyment* and very low in *Meaningfulness*. If scores were treated uni-dimensionally, the outcome would have been a neutral score on the evaluative dimension. Gutman and McConaughy build a compelling argument for discriminating between true indifference toward a program (where nothing is liked or disliked) and ambivalence (where a program is both liked and disliked). Their analysis of audience reactions to several popular programs indicated that over 90 percent of the respondents who assigned neutral scores to programs were ambivalent toward the programs. Based on the work of Kaplan (1972), the authors have devised a procedure for assessing the degree of liking and the degree of disliking a program. These two components are summed for each respondent to produce a Net Affect score. The Net Affect model is a better predictor of viewing (and in particular nonviewing) that the more traditional uni-dimensional model that does not consider ambivalent attitudes.

One of the major arguments against the use of a single evaluative scale is that it may favor a specific program format or other classification of programs. In their analysis of audience size and appreciation, Barwise, Ehrenberg and Goodhardt (1979) classified programs into one of two main groups: information and entertainment. Their findings indicate that information programs receive higher Appreciation scores than do entertainment programs. A similar trend was obtained in the qualitative ratings pretest conducted for the National Citizens Committee for Broadcasting (Philport and DeVerneil, 1978). That NCCB analysis was restricted to audience evaluations of prime time programs on the commercial networks, almost all programs were classified as entertainment programs. However, particular formats tended to get higher qualitative ratings. Family dramas (such as *The Waltons*, *Little House on the Prairie* and *Lou Grant*) and situation comedies (such as *M*A*S*H* and *All in the Family*) tended to receive higher qualitative ratings than action adventure programs (such as *Starsky and Hutch* and *Charlie's Angels*).

Under the economic viewer choice models of Wiles (1963) and Owen, Manning and Beebe (1974), it is feasible that entertainment programs and action-adventure programs reflect second choice, third choice or mass appeal program types that do not provide the same level of audience appreciation as those program types that received higher appreciation or qualitative scores. However, there is no way of empirically testing the assumption of these models without comparing evaluation across a variety of evaluative scales. Bowman's (1972) TV Viewing Choice model suggests that preferences for program characteristics are unique for each audience member and cannot be classified by predetermined program types. Bowman's model is partially supported by the failure of behaviorally defined patterns of program

preferences and exposure patterns (for examples, see Ehrenberg, 1968 and Philport, 1975) to match structurally defined program format classifications.

III. MULTI-DIMENSIONAL MODELS OF PROGRAM EVALUATION

From the studies and models described above, it appears that individuals in the audience apply their own criteria for classifying programs by evaluating an array of program elements that cannot be measured in a uni-dimensional evaluative scale.

A. Himmelweit, Swift and Biberian

A study by Himmelweit, Swift and Biberian (1978) represents one of the most exhaustive examinations of audience attitudes to be conducted in recent years. It is one of the few investigations to collect cognitive and affective reactions to programs as well as evaluations of stylistic elements. To clarify the distinction between reactions and responses to style, several of the variables used in this analysis are listed below:

Reactions	Responses to Style
Informative	Violent
Exciting	Funny
Trivial	Sexy
Moving	Unpredictable
Convincing	Complex

Their results indicate several key findings that are relevant to the development of a multi-dimensional model.

- Audiences tend to have greater agreement in their evaluations of style and more variance in their reactions to programs.
- The liking of a program is more closely related with emotional or affective evaluations (i.e. exciting and absorbing) than with cognitive evaluations (informative, realistic and not trivial).
- The distinguishing characteristic of uniformity of reactions is program popularity. The more popular the program, the more uniform the reactions from its audience.
- As the impact of a program decreases, so does its rating (impact was operationalized from scales such as absorbing, moving, exciting and disturbing).
- The popularity of a program appears to be correlated with the richness and complexity of its stylistic content.

According to these results, the audience can serve as critic and evaluate the stylistic content of a program. The fact that stylistic response means (i.e. violent, funny or sexy) had smaller standard deviations than reaction means indicates that

the audience can set aside their predispositions and evaluate program content with a fairly high amount of agreement. Himmelweit, Swift and Biberian correlated stylistic responses and reactions with the audience's liking of programs. This procedure allowed the authors to assess the strength of each of these components of the multi-dimensional model in predicting the individual's overall satisfaction with a program. Their analyses demonstrated that stylistic responses are the weakest predictors of audience satisfaction.

The authors further state that affective evaluations (such as exciting and absorbing) are better predictors of program satisfaction than cognitive dimensions (such as realistic and informative). They noted that the impact of a program (absorbing, moving and exciting) was the most significant evaluative predictor of a program's popularity. The NCCB qualitative ratings pretest (Philport and DeVerneil, 1978) provides additional support of the significance of the cognitive dimension of program meaningfulness (or importance). Unlike the Himmelweit, Swift and Biberian analysis, the NCCB study collected evaluations immediately after viewing of the programs. The findings suggest that when evaluations are gathered immediately after viewing, a program's meaningfulness or relevance is the best predictor of the audience's overall satisfaction with the program. Although there are some semantic differences between two scales used in the two studies, they both support the importance of the evaluative component of the multi-dimensional model in documenting viewer satisfaction with programs.

One of the true advantages of combining audience reactions with evaluations of program style is the increased power of the multi-dimensional model as a diagnostic tool. While we recognize that it is important for producers and broadcasters to know that their programs are (dis)liked, it is more functional for them to know the direct relationship between reactions to specific elements of a program and viewer satisfaction. The quantification of audience reactions to style is not an easy task. Himmelweit, Swift and Biberian's research indicate that there is a direct relationship between the richness and complexity of stylistic elements in a program and the program's subsequent popularity.

B. Landis Model

The Landis Group (1979) has recently uncovered 14 factors of program evaluations. Most of the factors are comprised of questions designed to elicit reactions to stylistic content. However, as one examines the TV Factors, it is interesting to note the inclusion of the functions, and cognitive and affective reactions in the instrument.

TV DIMENSIONS AND FACTORS (LANDIS, 1979)

Dim. 1: What the Program Does for the Viewer

Factor 1—Provides Knowledge and Enrichment
Factor 2—Provides Diversion and Escape

Dim. 2: How the Program Affects the Viewer

Factor 3—Gives Fun and Amusement
Factor 4—Builds Tension and Excitement

Dim. 3: How the Viewer Feels About the Program

Factor 5—Positive Evaluation, wants to watch
Factor 6—Negative Evaluation, not interested

Dim. 4: The Program's Type of Appeal

Factor 7—Acceptable for Kid/Family Viewing
Factor 8—Primarily for Adult Entertainment

Dim. 5: The Content of the Program

Factor 9—Subject/People are Familiar and Realistic
Factor 10—Subject/People are Unfamiliar or Unusual

Dim. 6: The Viewer Relationship to the Program

Factor 11—Viewer becomes Emotionally Involved
Factor 12—Viewer is Interested Observer

Dim. 7: Sensory Emphasis of the Program

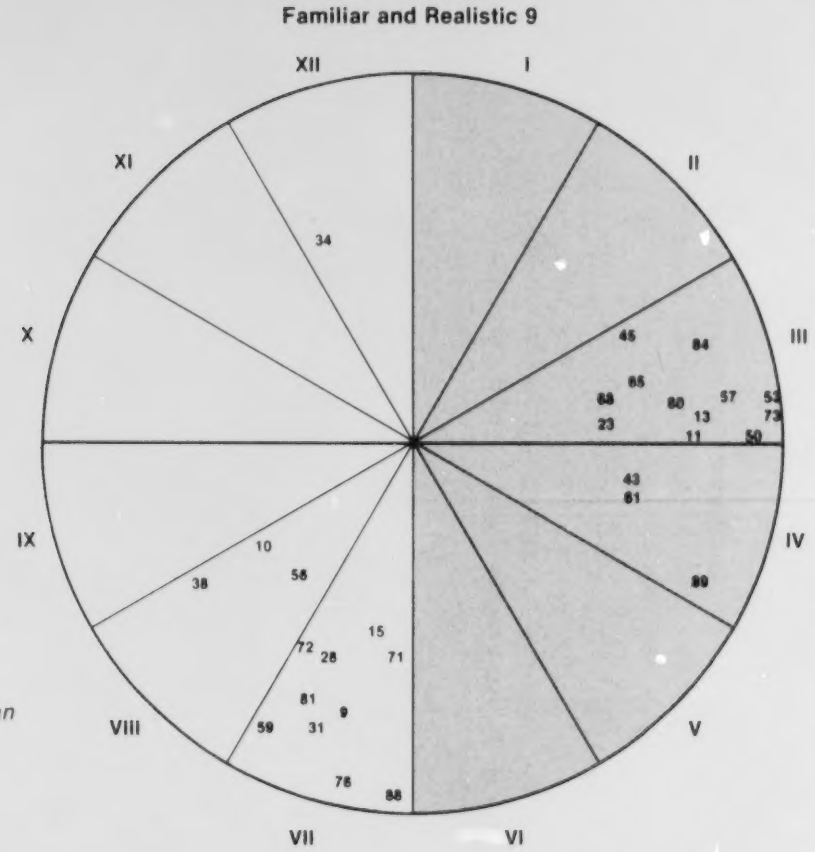
Factor 13—Visual Stress on Beauty and Glamour
Factor 14—Aural Stress on Wit and Clever Lines

At first glance the 14 factors do not illustrate the full scope of the instrument. Respondents were asked to evaluate television programs along 258 descriptors. Complex multivariate procedures were then used to "map" the evaluated programs along the 14 factors. It would not be within the purpose of this paper to present a complete discussion of procedures and results of the Landis analysis. However, two results are noteworthy. The first result is a confirmation of a finding in the Himmelweit, Swift and Biberian study—viewers can evaluate stylistic elements systematically and somewhat objectively. The second is an illustration of the interrelationships among program appeals and their impact in defining viewer discriminations among programs.

A graphic display of the relationships among television programs along two dimensions of the Landis study is presented in Figure B-1. Respondents perceived that *60 Minutes*, *McNeil/Lehrer* and *Meet the Press* were maximum "providers" of Knowledge and Enrichment, while *J. Cousteau* provided Knowledge and Enrichment toward the Unfamiliar and Unusual, and *Twilight Zone* presented Unfamiliar and Unusual Subjects and People. The comparisons could continue along the two dimensions in the graph,or combinations of other dimensions. However, from the perspective of the TV Viewing Choice Model (Bowman, 1975), it is important to note that Landis is one of the most recent researchers to document the viewer's ability to discriminate among programs based on their program appeal elements. In that the Landis instrument includes a *conative* dimension (Factor 5—Want to watch and

Figure B-1

An Illustrative Display of Selected Programs Along Two Dimensions of Program Style (Landis, 1979)



34 - *Happy Days*

2 Diversion and Escape

9 - *Battlestar Galactica*
10 - *Bewitched*
15 - *Captain Kangaroo*
28 - *Fantasy Island*
31 - *Gong Show*
38 - *Joe & Val*
56 - *Mork & Mindy*
59 - *Wonder Woman*
71 - *Sesame Street*
72 - *The Six Million Dollar Man*
76 - *Star Trek*
81 - *The Incredible Hulk*
88 - *The Twilight Zone*

23 - *Dick Cavett*
45 - *Lincoln Center*
50 - *MacNeil/Lehrer Report*
53 - *Meet The Press*
57 - *NBC News*
65 - *The Phil Donahue Show*
68 - *Roots*
73 - *60 Minutes*
80 - *TV Tests*
84 - *Today Show*
93 - *Wall Street Week*

Knowledge and Enrichment 1

11 - *Billy Graham*
43 - *Lifeline*
61 - *Nova*
89 - *Jacques Cousteau*

Factor 6—Not interested), analyses relating appeal elements with the intention to watch (not reported in Landis, 1979) could lead to stronger support of Bowman's model of viewer choice.

IV. THE PROGRAM EVALUATION PROCESS

Having reviewed the fundamental components of viewer attitudes toward television programs, it becomes apparent that many researchers have assumed that viewer evaluations of programs are fixed or static with little variation in attitudes from one episode of a program to another. A functional model of the television viewing process is presented in Figure B-2. It has been adapted from Rosengren's (1974) paradigm for examining "uses and gratifications" research. The "uses and gratifications" paradigm suggests that viewer attitudes toward a program can be conceptualized as either being associated with: one, a motive to view the program (or select another behavior), or two, a post-viewing evaluation of a program episode as providing a gratification or non-gratification. It is clear that most researchers have not identified this important distinction, although it may seriously impact the nature and range of reactions to the program.

As an illustration, the Himmelweit, Swift and Biberian study (1978) captured respondents' reactions to a limited number of programs listed on a roster. The list was biased in that it primarily included only popular programs. Most of these programs would have such strong images across the audience that unique individual reactions would be difficult to detect among the stereotypical reactions generated from discussions about the programs. In contrast to this approach, the NCCB study (Philport and DeVerneil, 1978) assessed immediate reactions after viewers had completed watching a program. This procedure fails to collect the attitudes of non-viewers who may have chosen not to watch due to specific negative attitudes toward the program.

Palmgreen and Rayburn (1979) have added an interesting approach by testing the relationship between gratifications sought and gratifications received from television viewing. The following model was used in their study:

$$E_i = \sum_{i=1}^{n} \frac{(GS_i - GO_i)}{n}$$

Where: E = an exposure for some medium, program content type, etc.

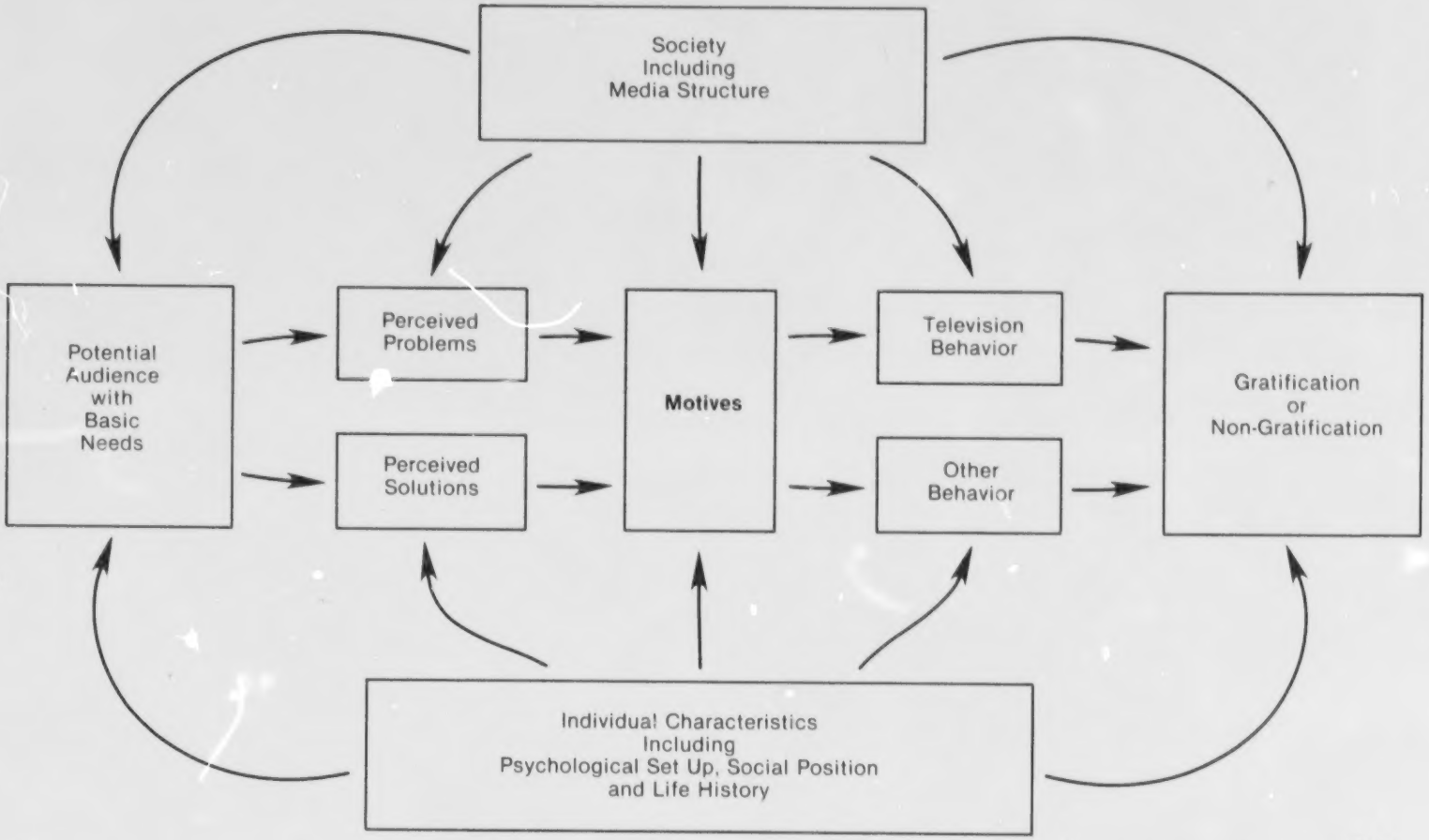
GS_i = a measure of the extent to which the ith gratification is sought.

GO_i = a measure of the extent to which the ith gratification is perceived to be obtained.

n = the number of gratifications under consideration.

Figure B-2

An Adaptation of The Rosengren Uses and Gratifications Model (1974) for Television Behavior



The authors used the gratification discrepancy model to predict viewers and non-viewers of Public Television. In the case of individuals who make their own television viewing decisions, the discrepancy model proved to be a significant predictor of Public Television viewing. In fact, it proved to be a better predictor than education, income or the number of children in the household. This model, for testing the difference between viewer predispositions and subsequent program evaluations, could also be employed to measure the degree to which a program meets the expectations of its audience. Any of the attitudinal components reviewed in this paper could be incorporated in the discrepancy model.

Intervening Variables and Audience Segmentation

Psychological needs are an important component affecting the media selection and evaluation process (McGuire, 1974). In addition to the more abstract psychological needs, marketing researchers have defined target audiences by their lifestyles (Plummer, 1974). A recent study by Frank and Greenerg (1979) has analyzed the relationship between leisure interests, psychological needs and television. The preliminary findings indicate that television viewers can be divided into specific interest groups with their own psychological needs. As audiences are segmented by psychological attributes, their evaluations as well as exposure to programs should be noted. Such procedures would help to determine if particular types of audiences are more satisfied with programs than other types of audiences.

V. CONCLUDING COMMENTS

This paper was intended to provide a conceptual presentation of psychological models of viewer evaluations of television programs rather than a discussion of the psychological derivations of those evaluations. It was designed to examine the assumptions of several attitudinal models and their utility in capturing functional information from audiences. We now need to proceed and employ the attitudinal models in systematic and theoretically based measurement of audience evaluations of television programs.

We should extend most of our energies in the examination of the relationships among existing cognitive, affective and stylistic evaluative dimensions. Most of the research reviewed in this paper suggests that uni-dimensional models of program evaluation cannot adequately capture the complete range of audience evaluations across all programs. Tests should include the examination of within format and between format analyses to examine format specific biases. Other testing should examine discrepancies between audience predispositions and subsequent program evaluations and determine when audience attitudes should be captured.

SOURCES

Bagozzi, R., A. Tybout, C. Craig and B. Sternthal (1970) "The Construct Validity of the Tripartite Classification of Attitudes." *Journal of Marketing Research*. 16, 88-95.

Barwise, T., A. Ehrenberg and G. Goodhardt (1979) "Audience Appreciation and Audience Size." Aske Research Ltd. (Prepared for the Independent Broadcast Authority).

Berkowitz, L. (1972) *Social Psychology*. Glenview, Ill.: Scott, Foresman and Company.

Bowman, G. (1975) "Consumer Choice and Television." *Applied Economics*. 7, 175-184.

Corporation for Public Broadcasting (1978). *The Effects of Television on People's Lives*. Washington, D.C.: Corporation for Public Broadcasting.

Ehrenberg, A. (1968) "The Factor Analytic Search for Program Types." *Journal of Advertising*. 8, 55-63.

Fishbein, M. (1965) "A Consideration of Beliefs, Attitudes, and Their Relationship," in I.D. Steiner and M. Fishbein, eds., *Current Studies in Social Psychology*. New York: Holt, Rhinehart and Winston, 1965.

Frank, R. and M. Greenberg (1979) "Zooming in on TV Audiences." *Psychology Today*. October, 92-103.

Gutman, J. (1978) "Television Viewer Types: A Q Analysis." *Journal of Broadcasting*. 22, 505-516.

Gutman, J. and D. McConaughy (1978) "Ambivalance and Indifference in Preference for Television Programs." *Journal of Broadcasting*. 22, 373-384.

Himmelweit, H., B. Swift and M. Biberian (1980) "The Audience as Critic," in P. Tannenbaum, ed. *Entertainment Functions of Television*. New York: Lawrence Erlbaum Associates.

Jeffres, L. (1975) "Functions of Media Behaviors." *Communication Research*. 2, 137-162.

Kaplan, K. (1972) On the Ambivalence-Indifference Problem in Attitude Theory and Measurement: A Suggested Modification of the Semantic Differential Technique." *Psychological Bulletin*. 77, 361-371.

Katz, E. (1960) "A Functional Approach to the Study of Attitudes." *Public Opinion Quarterly*. 24, 163-204.

Katz, E., J. Blumler and M. Gurevitch (1973)"Utilization of Mass Communication by the Individual," in J. Blumler and E. Katz, eds., *The Uses of Mass Communications: Current Perspective on Gratifications Research*. Beverly Hills: Sage.

Katz, E., M. Gurevitch and H. Hass (1973) "On the Uses of Mass Media for Important Things." *American Sociological Review*. 38, 164-181.

Landis Group (1979) *Mapping Analysis for Corporation for Public Broadcasting*. The Landis Group.

McGuire, W. (1973) "Psychological Motives and Communication Gratification," in J. Blumler and E. Katz, eds., *The Uses of Mass Communications: Current Perspectives in Gratification Research*. Beverly Hills: Sage.

Noll, R., M. Peck and J. McGowan (1973) *Economic Aspects of Television Regulation*. Washington, D.C.: Brookings Institution.

Osgood, C., G. Suci and P. Tannenbaum (1957) *The Measurement of Meaning*. Urbana: University of Illinois Press.

Owen, B., J. Beebe and W. Manning (1974) *Television Economics*. Lexington: Lexington Books.

Philport, J. (1975) "A Multivariate Field Study of Patterns of Television Program Exposure; Gross Consumption of the Mass Media; and Machiavellianism, Anomia and Self-Esteem," Unpublished doctoral dissertation, Bowling Green State University.

Philport, J., and A. DeVerneil (1978) *Pretest of the NCCB Qualitative Ratings Instrument*. Prepared for the National Citizens Committee for Broadcasting.

Rosenberg, M. (1956) "Cognitive Structure and Attitudinal Affect," *Journal of Abnormal and Social Psychology*. 53, 367.

Rosengren, K. (1973) "Uses and Gratifications: A Paradigm Outlined," in J. Blumler and E. Katz, eds., *The Uses of Mass Communications: Current Perspectives in Gratification Research*. Beverly Hills: Sage.

Steiner, G. (1963) *The People Look at Television*. New York: Alfred Knopf.

Steiner, P. (1952) "Program Patterns and Preferences, and the Workability of Competition in Radio Broadcasting," *Quarterly Journal of Economics*. 66.

Wicker, A. (1969) "Attitudes versus Actions: The Relationship of Verbal and Overt Behavioral Response to Attitude Objects," *Journal of Social Issues*. 25, 41-77.

Wiles, P. (1963) "Pilkington and the Theory of Value," *Economic Journal*. 73.

APPENDIX C
METHODOLOGICAL ASSESSMENT OF EXISTING EVALUATIVE RATINGS SYSTEMS

James A. Anderson

University of Utah

Department of Communication

AUTHOR'S PREFACE

The short lead time for this paper precluded any opportunity to provide a source supported review paper on the subject of evaluation measures. No loss, as a paper of that sort would undoubtedly be inappropriate to the select audience of this conference. What this paper attempts to do is to raise the issues which surround a major commitment to the use of the proposed measures. The issues have not been treated delicately nor with the typical academic appearance of objectivity. While not wishing to incite, I have attempted to delineate these issues with the crispness that one actually finds in the world of applications. The opinions expressed are my own.

The first part of this paper repeats the overview of the qualitative measures provided in the pre-conference material. My thanks to Carol Keegan and the OCR staff. The discussion that follows assumes the reader's familiarity with these services and on occasion references them by number.

Other than this Preface I have not used the word "qualitative" to refer to the measures but rather the word "evaluative." With some restraint, I will simply comment that the measures proposed in the conference come from a deductive and reductionistic perspective. There is a whole branch of science which is neither. The word "qualitative," if it has any meaning at all in research, refers to the methods used by the phenomenological branches of social science and not to numerically based questionnaires or data collection devices. A small point, of course, but one on which we turn.

BY WAY OF REVIEW

EXISTING EVALUATIVE TELEVISION RATING SYSTEMS AND RELATED RESEARCH METHODS

---

1. System: Percy System

Testing Site: viewers' homes

Respondents: volunteers, 200-household panel; all permanent residents over 5 years of age.

Materials Evaluated: TV commercials and programs evaluated while panelists view; viewer pushes buttons with evaluative labels: "Excellent," "Zap," "Dumb," "Informative," "Credible," "Funny," "Unbelievable," "Boring"; discrete (vs. continuous) responses recorded whenever the viewer chooses; two additional buttons, "Identification" and "Personality," are used to identify the specific household panelist evaluating programming and to apply any evaluative response specifically to an actor, character or other personality.

Data Collection Procedures: ten-button special response device for online recording of evaluations of viewed material; data transmitted to central computer via special telephone line.

Decision-making Uses of Data: identify copy points and program elements related to viewer assessments of appeal and credibility; input to commercial copywriting and to scheduling decisions; pilot testing of television programs; analysis of audience flow; qualitative ratings of current programming.

Limitations: limited sample; possible panel effects; lack of overall program evaluation index; appropriateness of button labels undetermined.

---

2. System: TvQ

Testing Site: Home

Respondents: c. 1,000-viewers drawn from a national household panel, Household Testing Institute (HTI).

Materials Evaluated: over 300 current CTV programs per survey (7 surveys a year).

Type of Data Collected: self-reported program familiarity and "likeability"; 1-5 rating scale evaluation of overall program (dis)satisfaction; respondent demographics; viewing predispositions, given hypothetical counterprogramming.

Data Collection Procedures: self-administered mail questionnaire.

Decision-making Uses of Data: likeability/familiarity ratio; scheduling for audience flow; counterprogramming; analysis of target group program preferences.

Limitations: possible panel effects; poor minority group representation; restriction to overall program evaluation ratings prevents diagnostic uses of data.

---

3. System: PTvQ

Testing Site: viewers' homes

Respondents: rotating sample of c. 1200 PTV viewers drawn from a national HTI Panel; oversampling of Blacks and Hispanics.

Materials Evaluated: c. 100 PTV programs; TvQ evaluations of commercial programs by the same respondents are also collected.

Types of Data Collected: same program familiarity question and 1-5 rating scale as are used in TvQ; amount of PTV viewing; respondent demographics.

Data Collection and Analysis Procedures: self-administered mail questionnaire; as with TvQ, two major indices reported:

- Familiarity Index: percentage of all respondents claiming familiarity with a program.
- Q Score: percentage of all respondents familiar with a program who say it's "one of my favorites."

Decision-making Uses of Data: same as for TvQ; additionally, trends public awareness of individual PTV programs, and can be used for formative input to develop targeted programming.

Limitations: vague measure of program familiarity; overall program evaluations do not allow diagnostic applications of the data.

---

4. System: TQR

Testing Site: home

Respondents: 2,300 viewers 12 years of age and older, drawn from HTI national panels.

Materials Evaluated: 168 current and past programs along 14 evaluative factors:

1. Knowledge and Enrichment	vs.	2. Diversion and Escape
3. Fun and Amusement	vs.	4. Tension and Exitement
5. Positive Evaluation	vs.	6. Negative Evaluation
7. Kid/Family Viewing	vs.	8. Adult Entertainment
9. Familiar and Realistic	vs.	10. Unfamiliar or Unusual
11. Emotionally Involved	vs.	12. Interested Observer
13. Beauty and Glamour	vs.	14. Wit and Clever Lines

Data Collection Procedures: self-administered mail questionnaire

Uses of Data: evaluate new program concepts; monitor established programs; identify programming gaps; guide scheduling decisions; develop program typologies; identify specific criteria for pilot testing

Limitations: may inundate decision-makers with data; stability of identified factors over time yet to be measured

---

5. CBS Viewer Session Room
Testing Site: special viewing room
Respondents: small sample recruited per test session
Materials Evaluated: network programs and pilots
Type of Data collected: (dis)liking of programs; probability of viewing specific programs, given hypothetical counterprogramming.
Data Collection Procedures: response device for online recording of (dis)liking of viewing programs; supplementary questionnaires and personal interviews
Decision-making Uses of Data: to identify high points and low points of audience appeal time-locked with programs evaluated; program concept testing; ratings diagnosis
Limitations: small sample; sample selection bias; artificial viewing environment

---

6. System: Audience Studies, Inc. (ASI)
Testing Site: special theatre
Respondents: c. 400 per session: recruited "off the street."
Materials Evaluated: radio and television commercials; feature films; TV programs.
Type of Data Collected: (dis)liking of viewed material; viewer demographics; TV usage; consumer behavior; recall of viewed material.
Data Collection Procedures: rheostat dial for online recording of evaluations of viewed materials; supplementary questionnaires; small group discussions (recorded).
Decision-making Uses of Data: predicted audience share against "average" program competition; program concept testing.
Limitations: sample selection bias; viewer samples not controlled for representativeness; artificial viewing environment.

---

7. Method: Warner Two-Way Cable: QUBE
Testing Site: home
Respondents: c. 100,000 subscribers in Columbus, Ohio.
Materials Evaluated: television program content.
Type of Data Collected: public opinion data, marketing research information, interactions with live programming.
Data Collection Procedures: special response buttons on the channel selector device; special channels available for narrowcasting.
Decision-making Uses of Data: direct marketing; program concept testing; pilot tests.
Limitations: limited sample; sample self-selection bias; possible panel effects.

---

8. System: Program Evaluation Analysis Computer (PEAC)
Testing Site: any location; system utilizes portable response device

Types of Data Collected: second-by-second viewer reactions to television programs; variable labels for 16 response keys; respondent demographics

Respondents: usually a group of individuals gathered to evaluate programming material; viewers, however, can also be geographically distributed (e.g., in their own homes).

Data Collection and Analysis Procedures: viewer reactions to program content; data displayed with video color graphics and traditional statistical formats; also provides hard copy data reports.

Decision-making Uses of Data: group testing of television material at various stages in the production process; usually used for formative research purposes; used to test effectiveness and appeal of educational programming for various audience subgroups; also might be used for post-broadcast evaluation of programs.

Limitations: limited memory capacity of response devices; unsupervised in-home use of devices by viewer panels yet to be tested.

---

9. Method: Uses and Gratifications Research

Testing Site: home or special research site

Respondents: sample selection criteria vary.

Information Desired: TV viewing motivations (implicit and explicit); viewer expectations of program as derived from uses, and as they influence gratifications.

Type of Data Collected: psychological and sociological "needs"; viewer expectations; TV consumption; viewer gratifications.

Data Collection Procedures: self-administered questionnaire or personal interview.

Decision-making Uses and Limitations of Data: *rarely* intended for, or directly useful in program scheduling or production decisions.

---

10. Method: Viewer-Defined Program Typing

Testing Site: home or special research site

Respondents: sample selection criteria vary.

Information Desired: dimensions along with viewers compare programs, independent of producer-defined "program types."

Type of Data Collected: perceived similarities and differences among programs; overall program preference judgments.

Data Collection and Analysis Procedures: questionnaires and personal interviews; repertory grid, multi-dimensional scaling, factor analysis, cluster analysis.

Decision-making Uses of Data: identify new program concepts; counterprogramming; programming gap analysis.

Limitations: applied uses of research results yet to be tested.

I. SAMPLING

In the ever-evolving social sciences, one can see the ever so slight erosion of the mythic position that sampling holds as the touchstone of science. It is now permissible to place on the stand the assumptions that sampling itself makes about the phenomena under study. This section briefly considers those assumptions when sampling people, behavior and events, and then moves to a discussion of the issue of representativeness. They follow in order.

Assumptions of Sampling

The primary assumptions of all sampling are consistency and uniformity. These terms have slightly different implications within each of the major forms of sampling which are appropriate to this conference: people sampling, behavior sampling and event sampling. We will consider each of these sets.

People Sampling

Although the direct interest of this conference is behavior—attraction, information gain, attitude change—for pragmatic and political reasons most often our sampling is of individuals. Pragmatically, individuals provide a handy repository of behaviors and a large number of behaviors can be collected at the same time. Politically, individuals, not behaviors, hold power which, of course, is no small concern for a public corporation. The sampling of people typically uses a demographic sampling frame. That is, individuals are selected to meet certain requirements of age, sex, race, socio-economic status and the like. This sample frame is used to capture the separate criterion variable—the behavior or behaviors under study. Given these circumstances, then consistency requires that the population sampled maintain stable characteristics over time in relation to the criterion under investigation. Uniformity means that each time a given characteristic appears in an element sampled it has the same conditional effect on the criterion. The theoretical import of these assumptions is that they permit reductionistic explanations. That is to say, the researcher can make meaningful statements about aggregate members of the population falling into these consistent and uniform characteristics, instead of having to make individual statements about each member of the population.

In the applications environment it means that the researcher can identify a characteristic called race of the respondent which with each individual of the population maintains a stable relationship to the criterion of, say, viewing choices, and that that relationship is uniform across all individuals so identified. Clearly a potent assumption and clearly one yet to be demonstrated. Media choice studies using *multiple* population characteristics rarely accommodate more than 20 per cent of the variance in the choice measure. Individual characteristics, then, show what can be described as significant but trivial relationships.[1]

Choice criteria have, as explored in Section III, particular problems when measured in the circumstances of everyday viewing. What of other criteria such as attraction, information gain, or attitude change? The literature from interpersonal

and audio-visual instruction studies using these variables is simply massive. It unquestionably demonstrates that these measures are not stable over individual expressions of content nor over classes of individuals. The assumptions of consistency and uniformity are not met.

Behavior Sampling

It is entirely possible, of course, that a single individual may over time emit all possible behaviors associated with the human condition and television. Behaviors, therefore, can be sampled from an individual or, in the interests of time, from a relatively small number of people. In behavioral sampling the notion of consistency implies that the preconditions of behavior are the same for each appearance and uniformity, that the consequences or purposes of the behavior will be the same.[2] Consistency and uniformity produce the notion of behavioral traits—the regularities which give definition to the concept of personality and personal reliability.

The most important implication, for our purposes, is that the content characteristics of the program, the element we can control, is a necessary and sufficient descriptor of the preconditions of the behavior we will measure (choice, attraction, information gain and/or attitude change). A second implication is that the consequences of that choice (attraction, information gain, or attitude change) will be congruent with the goals of the Corporation. Neither appears certain, an argument we will return to in Section III.

Event Sampling

Inherent in the questions asked by the measures proposed in this conference is the conditional relationship between viewing television and the behaviors of choice, attraction, information gain, attitude change. The event of viewing, then, is part of the sampling process along with individuals and/or behaviors. The event can be deliberately or inadvertently sampled. It is deliberately sampled when the event is permitted to take different values in the analysis. It is inadvertently sampled when the event is assumed to be constant. As with people sampling, the requirements of consistency and uniformity are met in event sampling when the population of events maintains stable characteristics over time in relation to the criterion and when each of those characteristics has the same conditional effect on the criterion. Inadvertent sampling appears to have little opportunity to meet these requirements.

Research in communication is just beginning to sketch what it is to "watch television." Observation studies[3] have shown television participation to have several expressions. Participation may be continuous or highly interrupted. Attention can be fully or partially connected. Content can be significant to the event or irrelevant. One may watch a program or just television. Viewers may be seeking information, fantasy, escape or whatever from the same program. Even viewing may be a misnomer, for the aural portion is often more dominant than the visual.[4] Participation may be continuous or highly interrupted.

It is, of course, much simpler to ignore these variations and to treat viewing as if by

an attentive and evaluating audience, as we would while taking continuous or coincident measures of attraction, information gain or attitude change. Our observation studies, however, indicate that this condition of viewing is less likely than all the others. What is the event that we measure when we collect data in this manner? Normal viewing—certainly not. Findings based on these collections can only report what *can* occur under artificial conditions, not what does occur.

Further, what is the event measured when we ask individuals to sum across all these variations to give a generalized evaluation of a series or personality? Is such an evaluation a meaningful part of watching television? The answer is unknown to this author.

Representativeness

Fledgling statisticians are taught to repeat by rote: "The only sample which can be representative is a random sample, but a random sample is not necessarily representative. A representative sample is one in which all the characteristics[5] of a population are present in the sample in the same relative proportion as in the population."

Unfortunately the researcher can never know that a sample is or is not representative unless the population is known (in which case one need not sample). A researcher can create arguments in support of the representativeness of the sample by showing its fidelity to known characteristics other than the criterion measure (e.g., people sampling and demographic characteristics). The weakness of this argument is that the relationship between these marker variables and the criterion is always unknown.[6] Consequently the value of the argument is unknown. For example, if sex of the respondent is unrelated to attraction scores of TV programs in the population, then showing a male/female split in accordance with census data is irrelevant. The argument is further weakened when we consider that the relationship between an evaluative measure of a television program and any demographic variable undoubtedly changes as content changes.

The argument of representativeness is essentially an argument about appearances. Representativeness is an issue in its own right only when the sample will be used to predict a *knowable* population event. The representativeness of the Gallup or Harris poll is an issue with an exact solution when measuring voting behavior; it is an issue with no solution when measuring public opinion.[7]

In the evaluative measurement of television program audiences, representation would be a defineable issue if we were interested in predicting audience size. In this case, of course, we would not be predicting the viewing choices of the population of TV viewers but that of the clearly biased sample used by the rating companies.

Beyond that specific instance, representativeness is an issue only insofar as the appearance of representativeness increases the value of the data collection for some end. Very simply, then, there is no method which will guarantee a representative sample, nor any method to demonstrate the representativeness (or the lack thereof) of the sample on some criterion. None of the methods used by any rating system will produce a random sample (all are dependent on the voluntary participation of the

respondents), but a random sample is the only sample with the potential for representativeness. The argument from and about representativeness is, therefore, an empty argument.

The argument is: Does this sample have the characteristics necessary to accommodate the persuasive requirements of this research? Sampling characteristics will rightfully deviate from the traditional canons of representativeness when the research purposes dictate. Research, after all, is an *argument* for the construction of reality not its validation.

II. TELEVISION PARTICIPATION AS A PROCESS

The topic of this section was presaged in the previous one under the heading of behavior sampling. There we reviewed the notion of behavioral traits—those regularities of behavior which permit us to speculate on conditional relationships, the causes and motives for effects. We became concerned because the implied perspective of evaluative measures is that the findings of those devices related to attraction, information gain and attitude change are the consequence, at least in part, of the content of television programs. We are all aware that consequential relationships impose the two conditions of necessity and sufficiency on the antecedent conditions. In the case of our application these requirements mean that the manipulation of program content or its scheduling and distribution is the sole explanation needed for the subsequent measurement results. Under these conditions management decisions are relatively clear-cut. Objectives are set based on measurement values. Manipulations, even those of trial and error, are attempted until the product returns the required results.

Unfortunately, difficulties compound rapidly in this Aristotelian world when: (1) competing explanations can be introduced for the results found in the analysis of a *system* of sufficient but not necessary antecedent conditions, and (2) that system operates in the manner of a *process* with equifinality of outcomes.

Television Participation as a System of Variables

If the reader will indulge me, I have chosen to use the word "participation" as a substitute for "viewing" to cover the variety of activities which actually occur vis-a-vis television use. Participation with television programs appears to occur within a system of variables. This system has at least four elements: the production team, the distribution environment, cognitive states, uses and gratifications of the audience, and the reception conditions. These elements are part of a symbiotic system in which changes in one affect all. A change may be countered, delayed, accelerated or nullified by concomitant action within the system. Interresponsive systems limit the utility of linear causative explanations. Multiple explanations are possible. Success by a student on a final exam might be the result of the content, its

presentation, the motivation of the student, the availability of the exam, prior knowledge of the student or any combination. Naturally each interest group will use whichever explanation is more profitable.

Data collections from systems of variables based on the conditional model always involve some control of the system. That is, as the independent variable is manipulated and the criterion is free to covary, the other elements of the system are held constant. This method, as earlier argued, permits one to describe what can happen under a given circumstance but not what does happen. Of course, this explanation may be sufficient for the purposes of the research. It may be enough to show that information gain can occur to justify the funding of a program.

The System in Process

If the system of variables is closed or static—the elements remain in a stable relationship—then the operation of the system is essentially predictable. Social systems, however, are not governed by the second law of thermodynamics and are therefore open—the elements are in a continual process of negotiating their relationships. These are systems in process whose outcomes are characterized by equifinality. Any outcome is as likely as any other at the start of the system. Television participation is clearly a system in process. Program styles wax and wane; the sitcom is fading and the western may return. Program material gains and loses acceptance; one can advertise feminine products but cannot tell ethnic jokes. Broadcast television competes with an ever-increasing array of distributions modes; the home recorder could make a shambles of the best designed audience flow schedule.

Open systems lessen the importance of individual studies as each individual study is only a snapshot of what has been. One will not learn what will be successful by considering the characteristics of *previous* programming nor from assuming certain constant audience types.[8] A study therefore, is never definitive. Only successive studies over time provide insight into the direction the processes may be taking. In every case, the extension of the knowledge gained into the future must be quite limited as the system itself is not constrained by past events. One must continually test the current to go with the flow.

III. MEASUREMENT ISSUES

This section explores the topics which can be raised in relation to the measurements under consideration at this conference. These measurements include those of attraction (like/dislike), information gain and attitude change. Inherent in the three is the basic measurement of exposure. We will take each of the first three measurements, discuss its characteristics, and consider its relationship and dependence on exposure.

ATTRACTION

Many of the evaluative measures reviewed for this paper (1, 2, 3, 4, 5, 6, 7, 8) make use of generalized notions of "like" and "dislike" (which I will classify as an attraction variable). The construct of attraction as it appears to be used in these measurements can be usefully divided into intrinsic and extrinsic elements.

Intrinsic Elements of Attraction

The intrinsic elements of attraction are those characteristics of content which draw audiences in and of themselves. An actor or a personality is perhaps the clearest example, but themes (violence, sex, action), characters (the cowboy) and settings all have attraction value for some segment of the audience. Clearly the operative attraction characteristics of given personalities, themes, and settings vary widely. The reality theme of a *60 Minutes* would be wholly inappropriate to the fantasy theme of a *Pink Panther*, but the themes of reality and escape are potent attractions per se.

Extrinsic Elements of Attraction

Extrinsic elements of attraction are the variable purposes and uses that viewers have for content. As these use demands change, whether in individuals or in the composition of the audience, the type of content which is the focus of attraction may also change.

The distinction between intrinsic and extrinsic elements of attraction has utility for program decision-making. In the first the program producer can make predictions of the attractiveness of a program by looking at content characteristics alone. That is, given a particular actor one can expect higher program sampling that given another although long run audience success is much less predictable.

With extrinsic elements the question of liking is dependent on the concomitant question of "for what?" For example, a viewer may well like the program *Connections* when the preconditions of need, mood and interest exist. Given a different set of pre-conditions the viewer may turn past that program to view *Three's Company*. Does this viewer now dislike *Connections*? An articulate viewer might respond: "I like *Connections*, but not tonight, dear." The attraction of *Connections* and *Three's Company* is, therefore, use dependent.

Attraction and Program Exposure

This season's rating history has demonstrated vividly, once again, the volatility of program choice. To examine the relationship between attraction and program choice, we need to hold in mind that attraction can have both absolute and conditional values and that content characteristics are not sufficient predictors of choice.

Our previous discussion of intrinsic and extrinsic elements of attraction deals with the first of these. As to the second, our observational studies and factor analytic studies conducted by McCain[9] have shown that some 50 per cent of television program choice is determined by factors other than the chosen program's specific

content. These factors include: the psychic and physical states of the viewer; the social, temporal, and proxemical conditions of viewing; technical aspects of signal reception and receiver operation.

By way of illustration, in many of our observed families control of the set passed among family members according to the time of day. A young child may have selection control in the early afternoon and pass that control to an older sibling as prime time begins, who in turn passes it to a parent in the later hours. Those who view, view what has been chosen for them. The decision for those not in current control is whether or not to watch television, not what program to watch.

Perhaps an easy way to show the constructed relationship between attraction and program choice is in a simple figure (Figure C-1) which has the chosen program in a small circle surrounded by a larger circle of choice determinates. This larger circle would first be divided in half. The upper half would contain the exogenous or content-free determinates of choice. The lower half would have the endogenous or content-bound determinates. This content-bound half would again be divided into the intrinsic and extrinsic elements.

A substantial research question is where to draw that division. The more we assign to the intrinsic arc, the greater the stability of the relationship between program choice and attraction (and, of course, any other measure which is content dependent). Unfortunately for the potency of content-based predictions, observation studies again show that while younger children can make space and time available for viewing particular programs with some ease, both the ability and interest of adults fall precipitously from adolescence through middle age.

The relationship between content and program choice is necessarily limited to those choice determinates which are content dependent. Those determinates I have suggested accommodate approximately 50 per cent of the variability of choice. The predictive power of any given content is dependent on the stability of the attraction value of that content. As that value changes so will its predictiveness. For individuals, only intrinsic elements show the necessary stability over time. For audiences, both the intrinsic and extrinsic elements can show this stability assuming that uses and gratifications have some cultural community. Our conclusion, therefore, is that content attractiveness is a limited but perhaps significant predictor of choice.

Program Exposure and Attraction

The obverse of the relationship between these two variables is also a useful study. What is the effect of increased exposure (for whatever reasons) on the attractiveness of programming? We have a number of "folklore" statements in television about the "big break" on the one end leading to the consequences of "overexposure" on the other. Further, we can ask what are the consequences of artificial exposure conditions of certain data collection procedures reviewed (1,5,6,8,). I do not pretend answers to these questions, although conceptually I cannot assume that the results from an assembled audience would correspond to the results from an audience viewing by choice.

Figure C-1
Choice Determinates

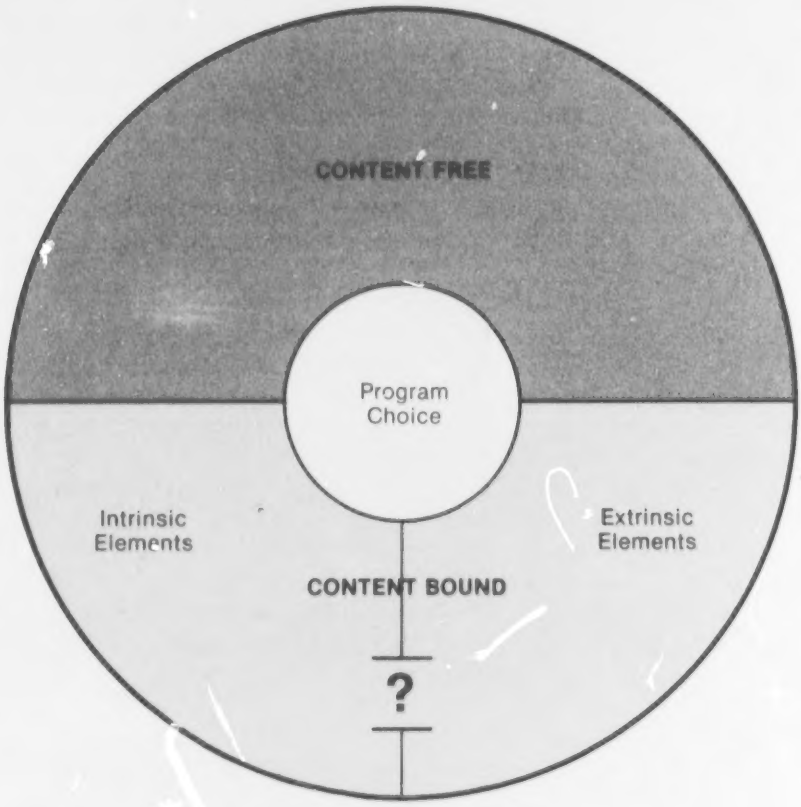


Figure C-1-- Projected distribution of choice determinates in television program selection. Content-free determinates are those operative when watching any television, regardless of what the program is. Content-bound determinates are of two classes: (1) intrinsic--the attraction is inherent in the content (2) extrinsic--the attraction is dependent on the utility the content has for the viewer. The relative proportion of intrinsic and extrinsic elements is variable and unknown.

INFORMATION GAIN

We suddenly take an exponential jump from the simple bifurcated measure of like/dislike to a massive array of possible measures and variables. As I look over a file of abstracts numbering in the hundreds, my recourse is to greatly simplify by a series of direct statements:

Information gain is most usefully approached as a learner variable rather than as a content variable. No technique of content presentation can guarantee information gain. Nevertheless there is a strong learner/mode of presentation interaction which is highly variable and dependent on the subject matter and the subsequent form under which information gain is measured.

Information gain has characteristics which interact with the mode of presentation, and the effectiveness of the learner. Some of these characteristics are duration, accessibility, transfer to other cognitive environments, and transfer to other modes of expression.

"Look," one may argue, "I want a kid to remember how to count to three. I test him beforehand and he can't. I show him a program and he counts to three. What's the problem?" (There is, of course, no problem as long as the researcher does not suggest that the child will be able to do it tomorrow, understands the notion of counting, can go home and count, or will be able to count in some cognitive task other than the one used in the test.)

Trying to earn my commission, I might ask in reply: "What if you ask two children and one counts and the other doesn't?" Are we then responsible for developing content so that both will count? And if the first revision shows the second child still unable to count, how many failures will we allow?

My position on this issue would be that the criterion of information gain places an excessive burden on the program agent. The research evidence from *Sesame Street* is simply too convincing of the frustrations likely.

ATTITUDE CHANGE

The concept of attitude change has shown a precipitous drop among communication scholars. There are many difficulties with the traditional definition of this concept and the measurement techniques used in its study. Traditionally attitudes have been defined as cognitive constructs which provide for consistency of behavior. One's attitudes direct one's choices. Attitude studies have shown that a single presentation with defined characteristics could change the attitudes of a group of college freshmen about tariffs on shoes. Unfortunately we have not been able to show results with issues of significance. Attitude change from the definition is a clinical concern, and best given to clinical methods.

More recent theorists have developed other ways of looking at attitudes and behavior. These ways define attitudes as explanations which each of us use to justify the behavior that we produce. These explanations occur after the behavior and are not presumed to predict it. It is a common event in our culture to be called upon to

give an explanation for behavior the actors themselves may not understand.

"Why do you smoke?"

"For the taste."

"It causes cancer."

"What does that have to do with taste?"

The attitude as explanation provides justification for one's actions and insulation from counter-arguments. Media presentations can be useful in providing the stuff from which these explanations can be constructed. Fewer and fewer scholars and researchers, however, espouse the theory of direct effects resulting in significant behavioral changes. Regularized behaviors are notable because they are *not* susceptible to rapid change; they are better considered part of articulated systems of mutual dependencies.

The notion of attitude change as a basis for program construction, selection or scheduling presents burdens at least of the weight of information gain. If the issue is trivial, why bother; if the issue is significant there is much more likely to be resistance to change and possible opposition to the message and messenger.

IV. SUMMARY OF METHODOLOGICAL CONCERNS

As way of introduction to the next section which considers the problems inherent in the pragmatics of using the proposed measures, this section briefly summarizes the concerns raised from the assumptions upon which these practices are based. The methodological concerns that we have discussed are: sampling, systems in process and measurement.

SAMPLING

We first examined the constructs underlying the three elements which are sampled in typical studies using the proposed measures: individuals, behaviors and events. With individuals we noted that the traditional demographic variables under which we classify individuals have been but marginally useful in providing explanations for viewing choice and program evaluations. Such studies assume a constant and uniform effect of program content on variables like age, sex of the respondent and race. It seems less than egalitarian to suggest that all persons over 54 have approximately the same taste in TV programming; or that all those interested in classical studies will enjoy British costume drama.

The behavior of these individuals is probably less a consequence of their position on some taxonomy than of the content of the programs that are presented. But even program content does not appear to be an adequate base for prediction; the uses and gratifications of the audience members for this content must play a significant part.

The event of viewing, with its multiple characteristics, is an element in this interaction also. We are just now approaching descriptions of what this event is.

Representativeness was the next topic. The point was made that it is much more an issue in the packaging of a presentation than in the methodologies available to program evaluations. Although appearances can be manufactured, no evidence of representativeness of the criterion can be presented without population knowledge of this criterion which of course obviates sampling. In a later section we noted that costly struggles for the grail of representativeness may lessen our actual knowledge of process events for, even if achieved at great expense, effort and revelation, the single sample is trivial in open systems. Better to spend one's time and money on successive sampling and interpretation over time.

SYSTEMS IN PROCESS

Television participation is a steadily evolving process with a continual infusion of new elements of judgment, applications, and industry practices. How little of the media of the 50s are comparable to the media of the 80s. A system in process comes from the past but is not bound by it. Evaluative research to be most useful will be a day by day, season by season activity.

MEASUREMENT

Under this heading, we first examined some of the characteristics of the judgmental or cognitive processes which are involved in presenting an evaluative response. We determined that only a relatively small percentage of the evaluative judgments were dependent on program characteristics alone. That is, there are few programs for which individuals have a general abhorrence or attraction. Television participation is not a highly passionate activity. Most programming is available to most individuals for some use. It is the uses that persons have or the gratifications they seek that interact with content characteristics to classify programs into utilities for some end. For example, reportedly, the highest rated program annually is the Superbowl. More times than not the game itself is boring. The utility of the Superbowl is not the game but the social constructs which have developed around the event.

Audience uses and gratifications are not constrained to the intent of program producers. That a program might be classified as informative does not prevent it from being evaluated as entertainment, fantasy, a social necessity or what have you by a respondent. The meaning of the evaluation, therefore, is contained in the respondent and not in the response.

The methods used to collect information, obviously, constrain the viewing patterns or construct evaluations which are not ordinarily used by the viewing audience. The opportunity to evaluate is an event in its own right separate from that of just viewing. An individual chooses to participate or not to participate.[10]

Global judgments about personalities, programs and series can gloss the complexities of the interaction among content, uses and gratifications and the preconditions of viewing. Consequently they will not predict viewing of, or satisfaction with a program.[11] They will not predict program concept or personality success by

looking at other programs. They are not a sufficient explanation for why one views although they *may* be part of one.

They do provide a definition of program success other than attendance and may consequently provide some insulation from circulation pressures. They may provide a source of justification for program development and programming decisions depending on their acceptance by funding agents and management. They have been used with some success to pre-test a program, program ideas and individual personalities.

V. PRAGMATICS OF USE

The assumptive bases are rich fields to mine for issues of concern in doing any kind of research. Particular applications show other problems in the actual practices of data collection, analysis and interpretation.

COLLECTION PARADIGMS

Collection of evaluative measures from respondents in free viewing conditions will present substantial problems for large data collection efforts. Assuming that free viewing choice of public television programs will continue to average around five per cent, the original sample will have to be 20 times larger than the desired sample size. Vendors which offer fixed sample sizes in the low hundreds have no capacity to provide samples of PTV viewers larger than 50 respondents (1).

There are, of course, a number of techniques which can increase the proportion of PTV viewers. They do require specialized sampling which eliminates the general populace samples and panels used by a number of the evaluative services (2,4).

Controlled viewing conditions involving either out-of-home viewing or special incentives to view lay the data collection open to criticism of bias (1, 5, 6, 7, 8). As I have argued this criticism cannot be effectively answered and may limit the political utility of the study.

COLLECTION DEVICES

Hard wired collection devices place obvious limits on sample size, geography, respondent selection and the proxemics of the respondent household (1,7). Both hard wired and no wire devices (8) impose substantial cost limitations on the data collection. The number of coincidental respondents is directly tied to the number of available instruments.

The complexity of operation of even simple devices introduces error which increases as the age of the respondent decreases. Most children below the fourth grade level will need direct supervision in the use of any device or questionnaire.

ACCESSIBILITY OF DATA FROM VENDED RESEARCH

Substantial restrictions are placed on the utility of purchased services when direct

access to the raw data is not permitted or possible. In such cases, the customer is invariably forced to adapt to the questions the vendor's display is capable of answering. When direct accessibility is not available the adaptability and speed of response of the vendor becomes an issue of increased magnitude. Accessibility to the data and all transformations is a hallmark of a preferred service.

ACCEPTANCE OF THE MEASURES

Whether or not CPB can maintain a substantial effort in evaluative research depends entirely on the level of acceptance such measures can gain outside OCR. The history of the rating services is instructive in considering the potential of the evaluative measures.

Television rating services are a beneficial parasite of the broadcast industry. There is no question that, given the current source of income for U.S. broadcasting, the rating services would be immediately re-invented if they should magically disappear.

From any but a commercial point of view the methodology of the rating services is less than rigorous. On the average, half of the original telephone-biased sample is lost in each market. The diary is difficult to maintain, memory dependent and error prone. Diaries are edited into usable form by arbitrary rules which clearly distort the findings. When findings which are even marginally detrimental to the industry appear, "methodological difficulties" are also found. The practices used for the instant ratings are even further from the norm.

None of this matters of course as there is no interest among the winners to change the rules of the game.[12] These measures are accepted and used with surety.

Altruism aside, whatever evaluative methods are used, they will be successful if they can be properly promoted to funding agents, and if programming which shows the desired response characteristics also conforms to management notions of what programming should be. One suspects that where funding agents go, management hearts and minds will follow.

PROGRAM CONSEQUENCES OF EVALUATIVE MEASURES

This section speculates on the direction programming might take should evaluative measures, particularly those of attraction, become widespread. Probably the closest analogy is found in the current developments of local news. Most major stations subscribe to some research service. These services use evaluative techniques which include the classes of measures reviewed here. While their major emphasis has been on the news personality, content style and components have also been studied. The trend in personalities is toward the young, bright, sometimes witty individual. Human interest stories, live coverage, video inserts and personal interchange have increased. More programs have humorous social commentary components. A crusading omniscience (a'la *60 Minutes*) is becoming evident in promotional activity and creeping into presentations. But, perhaps most significant is the unrelenting sameness of major market news programs throughout the country.

These characteristics may be the elements the desired commercial audience seeks. Sampling from other audiences may see programming lean in other directions.

MEASUREMENT COMPETITION

Studies from mediated instruction and commercial message effectiveness suggest that program construction recommendations from the three classes of measurement—attraction, information gain, and attitude change—may be competitive. Humor, for example, a significant attraction agent, has been shown to interfere with information gain and attitude change. The goals of each measurement approach are not necessarily complementary with the others.

POLITICAL RAMIFICATIONS OF INFORMATION GAIN AND ATTITUDE CHANGE MEASURES

The goals represented by measures of information gain and attitude change are heavily value laden. The adoption of these approaches introduces the clear and present danger of a political or social response to those values.[13] The wording of specific items has become a source of confrontation in educational research. Parents and school boards with the right of access[14] have censured items and objected to programs of instruction when they have learned of what the item measures. The pressure of polarized social action groups may result in politicization of Corporate research activities.

FOOTNOTES

[1] James A. Anderson, *Mass Media Utilization by Urban and Non-Urban Respondents*, Athens, Ohio: Broadcast Research Center Report No. 1, 1971.

[2] These requirements hold whether the behavior is repeated by a single individual or independently produced by several individuals.

[3] The University of Utah is part of a consortium of Universities building a library of case studies documenting televiewing in family settings.

[4] Considerable adult and adolescent viewing is coupled with other visually dependent activity. The viewer follows the action by sound sampling the video when need arises.

[5] In most applications the researcher is actually little interested in representing all characteristics, but is essentially concerned with the representativeness of the criterion measure(s) which in our case are the credibility, attractiveness or what have you of television programs.

[6] Evidence from correlations within samples cannot be used to support representativeness unless those samples can be shown to be representative.

[7] Certainly the outcome of the 1980 presidential campaign can be predicted accurately with a sample composed entirely of women. Unless one wishes to argue that women cannot vote like men.

[8] I realize that we have 50 years of broadcast industry thinking invested in program traits and audience types. When we ask "what is the favorite program of women 18-49" we assume that something is femaleness and that 32 year age span is conditionally related to content characteristics. Barring ties, some program will be a favorite simply by collecting the aggregate. Such analyses tell us nothing about women of any age or of television programs. The shards of series which litter the film vaults every season are evidence enough.

[9] Thomas McCain, "Managing Slop in the ROPJ Model." Lecture given at the Public Television Institute, Denver, Colorado, 1979.

[10] Generally less than one half of the Percy sample, for example, evaluates a given program when no other incentive is provided.

[11] They may "postdict" viewing, since evaluation is clearly dependent on exposure.

[12] In fact, the only certain effect of better methods would be increased costs.

[13]Even *Sesame Street* has its detractors. I have heard it described by a distinguished academician as a program "designed to produce a nation of filing clerks."

[14] CPB Research using these measures will necessarily follow the HEW guidelines for use of human subjects either by law or expediency.